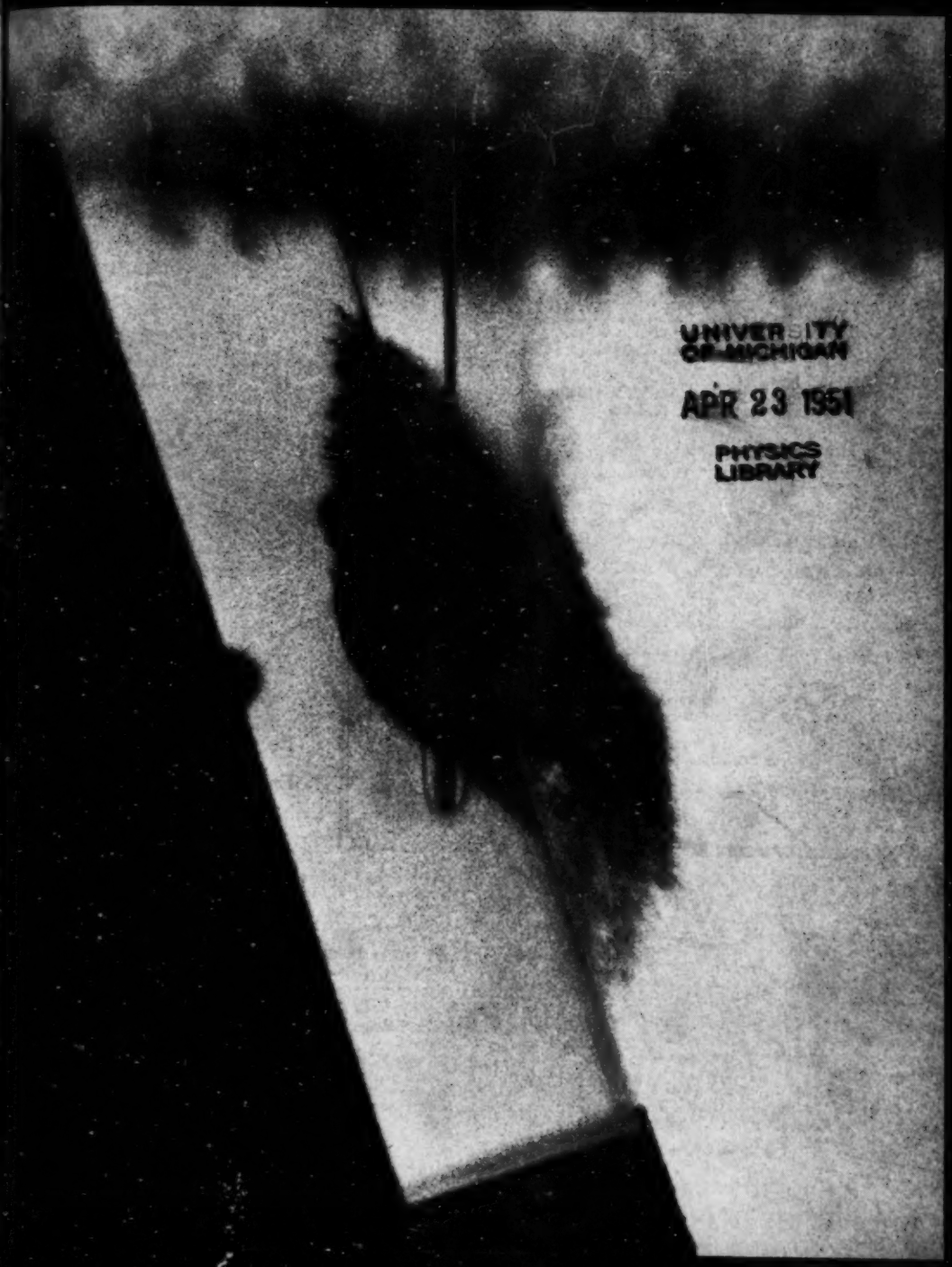
JOURNAL OF APPLIED PHYSICS



UNIVERSITY OF MICHIGAN
APR 23 1951
PHYSICS LIBRARY

JOURNAL OF APPLIED PHYSICS

Vol. 22, No. 4 APRIL, 1951

In This Issue

Page

Study of Precipitates of C and N in Iron with an Electron Microscope JOHN RADAVICH AND CHARLES WERT 367

Theory of D_0 for Atomic Diffusion in Metals CLARENCE ZENER 372

On the Scattering of Plane Waves by Soft Obstacles. I. Spherical Obstacles ROBERT W. HART AND ELLIOTT W. MONTROLL 376

Separation of Variables in Electromagnetic Theory DOMINA EBERLE SPENCER 386

Diffraction Errors in an Optical Measurement at Radio Wavelengths G. A. WOONTON, J. A. CARRUTHERS, H. A. ELLIOTT, AND E. CAROLINE RIGBY 390

Effect of Hydrostatic Pressure in an Electron Beam on the Operation of Traveling-Wave Devices PHILIP PARZEN AND LADISLAS GOLDSTEIN 398

On Stability of Linear Varying-Parameter Systems LOTFI A. ZADEH 402

Effect of a Grounded Slab on the Radiation from a Line Source C. T. TAI 405

Formulas for the Determination of Residual Stress in Wires by the Layer Removal Method W. T. READ, JR. 415

Directional Distribution of 1040-Kev Radiation from a High Voltage X-Ray Tube N. GOLDSTEIN, U. FANO, AND H. O. WYCKOFF 417

Lattice Distortion Spectrum of Evaporated Gold P. G. WILKINSON 419

High Temperature Structure and Thermal Expansion of Some Metals as Determined by X-Ray Diffraction Data. I. Platinum, Tantalum, Niobium, and Molybdenum JAMES W. EDWARDS, RUDOLPH SPEISER, AND HERRICK L. JOHNSTON 424

Oscillations of Magnetic Suspensions KNOX MILLSAPS AND J. C. MCPHERSON 429

Analysis of Synchronous Conditions in the Cylindrical Magnetron Space Charge H. W. WELCH, JR., AND W. G. DOW 433

Deformation of Gold Wire at High Temperature B. H. ALEXANDER, M. H. DAWSON, AND H. P. KLING 439

Amplification of the Traveling Wave Tube BERNARD FRIEDMAN 443

Interfacial Free Energy of Coherent Twin Boundaries in Copper R. L. FULLMAN 448

Crystallography and Interfacial Free Energy of Noncoherent Twin Boundaries in Copper R. L. FULLMAN 456

Leak Location by Radioactive Gases in Buried Pipes ANDREW GEMANT, EDWARD HINES, AND E. L. ALEXANDERSON 460

Diffraction Pattern of Microwaves Near Rods C. L. ANDREWS 465

On the Spectrum of Isotropic Temperature Fluctuations in an Isotropic Turbulence STANLEY CORRSIN 469

Lubrication at Extreme Pressures with Mineral Oil Films O. H. CLARK, W. W. WOODS, AND J. R. WHITE 474

X-Ray Measurement of Short Range Order in Ag-Au N. NORMAN AND B. E. WARREN 483

Kerr Cell Photography of High Speed Phenomena E. M. PUGH, R. v. HEINE-GELDERN, S. FONER, AND E. C. MUTSCHLER 487

Positive Coaxial Cylindrical Corona Discharges in Pure N_2, O_2, and Mixtures Thereof CHARLES G. MILLER AND LEONARD B. LOEB 494

Surface-Wave Propagation Over a Coated Plane Conductor STEPHEN S. ATTWOOD 504

Mathematical Basis of the Interpretation of Tracer Experiments in Closed Steady-State Systems C. W. SHEPPARD AND A. S. HOUSEHOLDER 510

Erratum: Diffraction Pattern in a Circular Aperture Measured in the Microwave Region C. L. ANDREWS 520

Electric and Magnetic Forces between Sphere and Wire W. R. SMYTHE 521

Announcements:

Symposium on "Instrumentation for Iron and Steel Industry" 371

RCA to Show New Electronic and Radio Techniques at 1951 Conference of Corrosion Engineers 414

Midwest Research Institute 428

Letters to the Editor:

Transmission Line Kerr Cell J. W. BEAMS AND H. S. MORTON, JR. 523

Temporary Enhancement of Hysteresis in Barium Titanate Samples D. R. YOUNG 523

Difference between Turbulence in a Two-Dimensional Fluid and in a Three-Dimensional Fluid T. D. LEE 524

Generality of Mixed Gas Flows MAX M. MUNK 524

Decay of Spectra Components in Isotropic Turbulence HIROSHI SATO 525

Propagation on Dielectric Coated Wires E. T. KORNHAUSER 525

Coefficients of Thermal Expansion of Au-Cd Alloys Containing 47.5 Atomic Percent Cd LO-CHING CHANG 525

Index to Advertisers xxxii

BOARD OF EDITORS

ELMER HUTCHISSON, *Editor*

EARLE C. GREGG, *Assistant Editor*

Associate Editors

HOWARD H. AIKEN
LYLE D. BORST
JAMES HILLIER
LEONARD B. LOEB
W. JAMES LYONS
ERIC REISSNER
RAYMOND J. SEEGER
FREDERICK SEITZ
CHARLES S. SMITH

Publication of the

American Institute of Physics

OFFICERS

GEORGE R. HARRISON, *Chairman*
GEORGE B. PEGRAM, *Treasurer*

GOVERNING BOARD

J. W. BEAMS
C. PAUL BONER
WALLACE R. BRODE
W. F. FAIR, JR.
S. A. GOUDSMIT
G. P. HARNWELL
GEORGE R. HARRISON
F. V. HUNT
RUDOLPH KINGSLAKE
PAUL KIRKPATRICK
F. W. LOOMIS
J. R. OPPENHEIMER
GEORGE B. PEGRAM
I. I. RABI
DUANE ROLLER
HAROLD K. SCHILLING
FREDERICK SEITZ
FRANCIS G. SLACK
JOHN C. STEINBERG

Consultant on Printing

MELVIN LOOS

ADMINISTRATIVE STAFF

HENRY A. BARTON, *Director*
WALLACE WATERFALL, *Executive Secretary*
THEODORE VORBURGER, *Advertising Manager and Assistant Treasurer*
RUTH F. BRYANS, *Publication Manager*
BETTY H. GOODFRIEND, *Assistant Secretary*
EILEEN B. NEUBERGER, *Circulation Manager*
KATHRYN SETZE *Senior Accountant*

The JOURNAL OF APPLIED PHYSICS, published monthly at Prince and Lemon Sts., Lancaster, Pa., is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as PHYSICS. The JOURNAL OF APPLIED PHYSICS publishes editorials and reviews, as well as technical papers of applied physics. Articles appearing in it are indexed regularly, according to their field, by Physics Abstracts, Chemical Abstracts, Engineering Index, Inc., Industrial Arts Index, etc.

Manuscripts should be submitted to Elmer Hutchisson, Editor, Case Institute of Technology, Cleveland 6, Ohio. Unless otherwise stated, submission of a manuscript is a representation that it has been neither copyrighted, published, nor submitted for publication elsewhere. The authors' institutions are requested to pay a **publication charge** of $8 per page which, if honored, entitles them to 100 reprints without further charge. Instructions will be sent with galley proofs.

Subscriptions, renewals, and orders for back numbers should be addressed to the American Institute of Physics, 57 East 55 Street, New York 22, New York.

Subscription Price	U.S. and Canada	Elsewhere
To members of the American Institute of Physics and Affiliated Societies	$10.00	$11.00
To all others	12.00	13.00

Back Number Prices

Yearly back number rate when complete year is available: $10.00. Single copies: $1.00 each.

Changes of address should be sent to the American Institute of Physics; **corrected proofs** should be sent to the Publication Manager, American Institute of Physics. **Advertising rates** supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879. Accepted for mailing at the special rate of postage provided for in paragraph (d-2), section 34.40, P. L. & R. of 1948, authorized May 2, 1932.

Journal
of
Applied Physics

Volume 22, Number 4 April, 1951

Study of Precipitates of C and N in Iron with an Electron Microscope

JOHN RADAVICH
Department of Physics, Purdue University, Lafayette, Indiana

AND

CHARLES WERT*
Institute for the Study of Metals, University of Chicago, Chicago, Illinois

(Received August 14, 1950)

By the use of the electron microscope it has been possible to study the precipitates of C and N in α-iron. The physical shapes of the precipitates have been determined in each case as well as the fineness of dispersion The results of the investigation confirm the conclusions drawn from internal friction measurements made on these alloys and reported earlier.

I. INTRODUCTION

IN the past three years there have appeared a number of papers by Dijkstra,[1] Wert,[2] and Zener[3] in which is described a series of investigations on the precipitation of C and N from supersaturated solid solution in iron. The main experimental part of this work made use of the internal friction which is found in materials with the crystal structure of these alloys. In these articles, a number of qualitative and quantitative statements about the precipitates formed have appeared.

1. In the precipitation of N under appropriate conditions two separate constituents form in succession. One of these is a metastable plate-like structure (of unknown chemical composition) which disappears when a second constituent (presumed to be Fe_4N) forms upon prolonged tempering.[1,2]

2. When C precipitates from dilute solution in α-iron, the precipitates (presumed to be Fe_3C) are formed in the shape of spheres. These spheres are of calculable size.[2]

Most of these claims have not been verified by any other common metallurgical technique, because the meager amount of precipitate formed and the high state of dispersion of the precipitates make such measurements extremely difficult. At the time that the above-mentioned papers were presented, it appeared that the most clear-cut way to prove or disprove the results of these admittedly indirect experiments was to try to observe the precipitates directly, using an electron microscope. The present paper shows the results of this investigation.

II. SAMPLE PREPARATION AND REPLICA TECHNIQUE

The specimens were flat strips of Puron iron 6 in. long, $\frac{1}{2}$ in. wide, and 0.50 in. thick. These samples were rolled to size, annealed, and decarburized before treatment. The supersaturated solutions were made by carburizing and nitriding the sample in hydrogen containing heptane and ammonia vapor, respectively. Tempering of the sample was then carried out; the progress of the precipitation was followed by means of electrical resistance measurements at -195 deg C.[4] All samples were tempered just long enough to precipitate the main part of the C or N, since excessive coalescence was not desired. The four specimens finally used for microscopic examination are designated in the following way:

I. Original material—no treatment after decarburization.
II. Iron +0.02 percent C. Precipitated at 300 deg C for 25 minutes.
III. Iron +0.05 percent N. Precipitated at 250 deg C for 30 minutes.
IV. Iron +0.035 percent N. Precipitated at 210 deg C for 6 minutes.

* Now at the University of Illinois, Urbana, Illinois.
[1] L. J. Dijkstra, J. Metals, Trans. **185**, 252 (1949).
[2] C. A. Wert, J. Appl. Phys. **20**, 943 (1949). C. A. Wert and C. Zener, J. Appl. Phys. **21**, 5 (1950).
[3] C. Zener, J. Appl. Phys. **20**, 950 (1949).
[4] C. Wert, *Thermodynamics in Physical Metallurgy*, A.S.M. publication (1949), page 178.

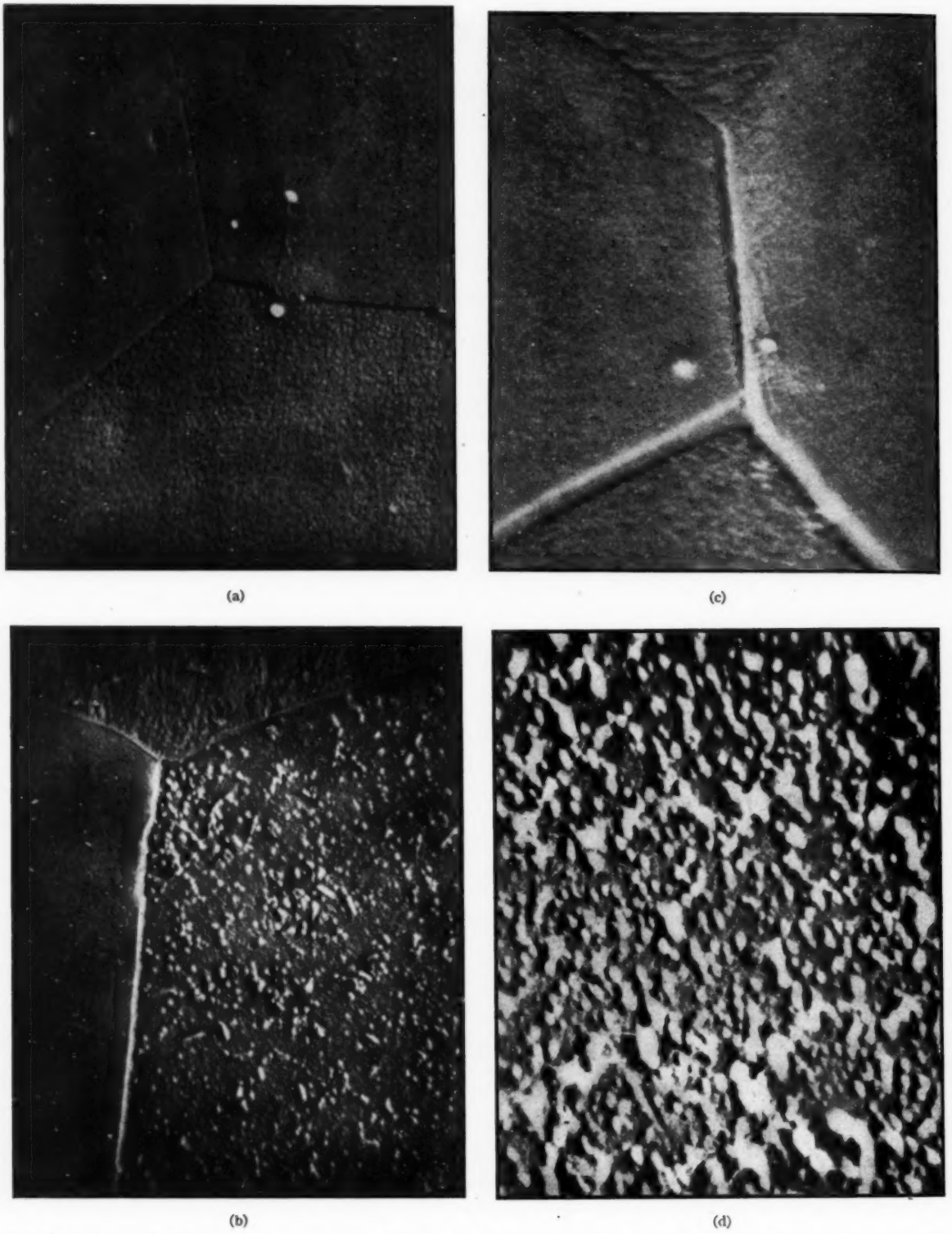
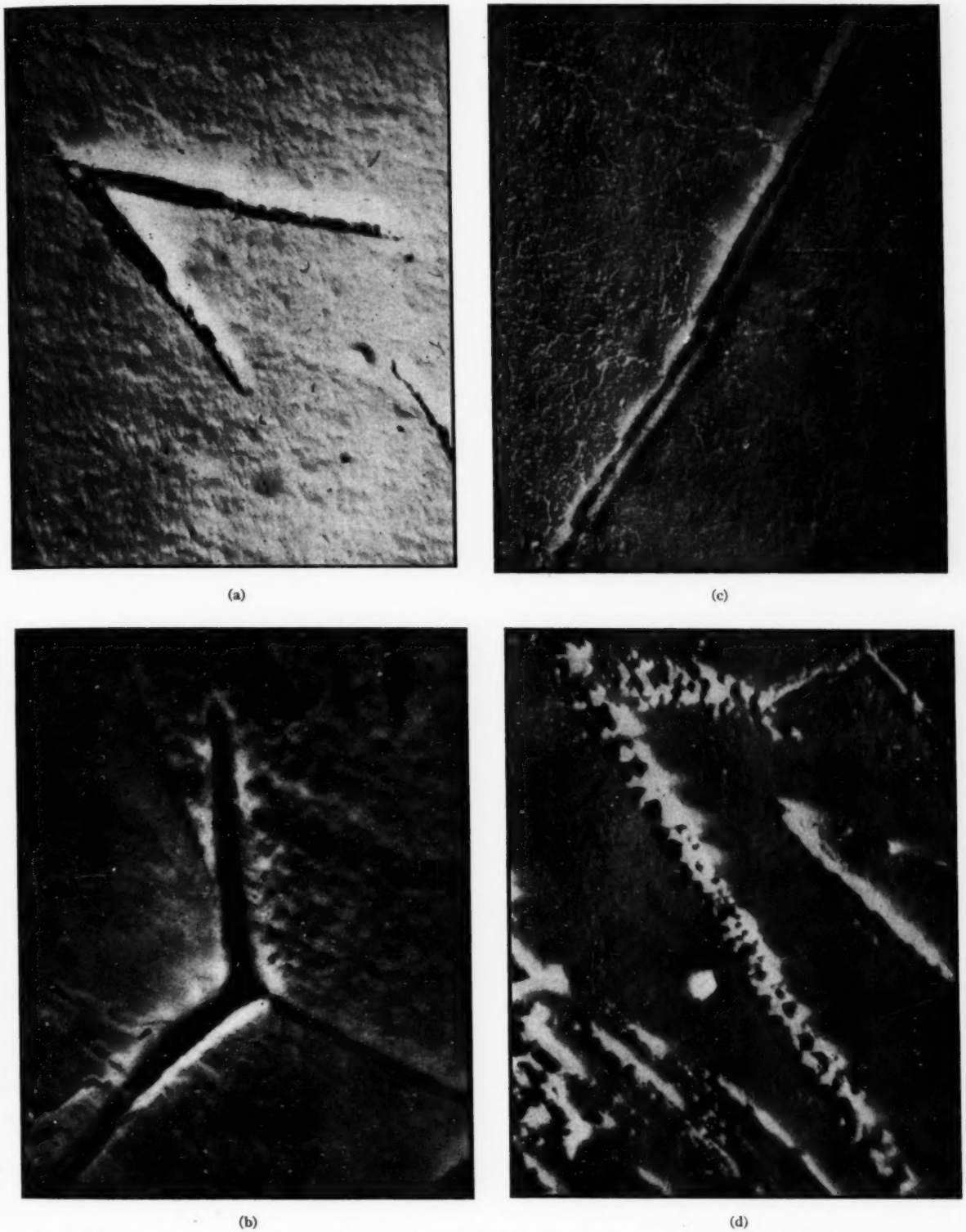


FIG. 1. (a) Formvar replica of original uncarburized surface after electropolish, specimen I. (b) Replica of surface of specimen II. (c) Replica of oxidized surface, specimen I. (d) Replica of oxidized surface, specimen II. Magnification ×16,400.

For examination, the specimens were cut into sections ¾ in. long and were rough polished through 3/0 paper. After washing and rinsing in alcohol, they were electropolished. This electropolish was carried out in a Buehler-Waisman electropolisher with an electrolyte of 850 cc of ethyl alcohol, 140 cc of distilled water, and 50 cc of perchloric acid. The polishing required 30 seconds with 30 volts and 0.9 ampere.

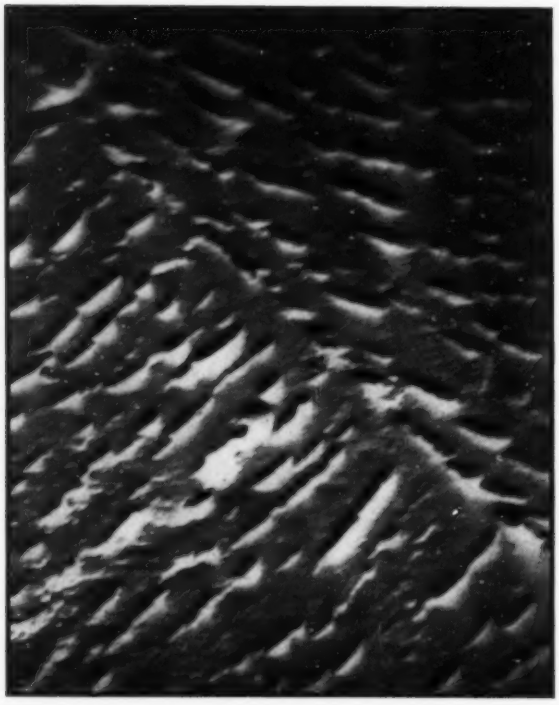


(a) (c)

(b) (d)

FIG. 2. (a) Formvar replica of electropolished section of specimen III. (b) Formvar replica of chemically etched section of specimen III. (c) Replica of oxide film formed on a section of specimen III. (d) Replica of a selected region of specimen IV as electropolished. Magnification ×16,400.

For the major part of the examination, a Formvar replica technique was used so that the structure of a single sample could be examined after repeated etchings. It was hoped that by this method an etching time would be found which was best suited to reveal the precipitate. It proved to be possible to produce the desired etch by



(a)

(b)

FIG. 3. (a) Replica of a section of specimen IV as electropolished. (b) Replica of a selected region of specimen IV as electropolished. Magnification ×16,400.

electropolish alone without the use of a chemical etchant. The replicas were shadow cast with chromium at a filament height to screen distance ratio of 1:7. This produced replicas of high contrast and clarity.

III. EXAMINATION OF THE CARBIDE PRECIPITATE

A sample from specimen I was first used for examination and aside from some very large inclusions (apparently an oxide), it shows no precipitates; Fig. 1a shows the structure present. The grain boundaries are free of any material and the grains have a smooth appearance. Certain grains are etched slightly by the electropolishing. All of the illustrations are prints made from original plates. The shadows are white.

Replicas of sections from specimen II were then prepared in the same manner; a typical one is shown in Fig. 1b. Examination of the replicas prior to shadow casting showed light spots representing thin regions in the Formvar replica; these represent raised regions on the surface of the metal. These small raised regions are presumed to be the carbide particles. The action of the electropolish causes the carbide particles to stand in relief; the particles themselves are not transferred to the replica but only their impression. The precipitate is most readily observed in those grains which have the greatest tendency to be slightly etched by the electropolish and is absent in the smooth grains. Large precipitate particles do not cluster in the grain boundaries, but rather there is usually a region some thousands of angstroms in width about a grain boundary in which no precipitate is found. This depleted region is presumably caused by migration of carbon to the grain boundary during tempering; any precipitate which forms there must be so finely divided that no structure can be detected with the resolving power obtained.

The impressions of the precipitate particles as shown in Fig. 1b are typical for all sections made in specimen II. As far as can be seen, the particles are indeed something like spherical as had been predicted.[2] Though they do not all appear to have the same size, the deviation in size seems remarkably small, especially since some of the particles have presumably not been made to stand fully in relief. Their size, as measured on the negatives, is found to be in the range of hundreds of angstroms, an average size would perhaps be some 500 angstroms. This measured size compares with a value of about 1500 angstroms predicted for this sample using the procedure of reference 2.

To provide a further check on the correctness of the use of this Formvar replica technique, a method using oxide films was also devised for these specimens. This method consists in making a thin oxide film on the surface of the metal, then using this film for examination in the microscope. It has been found from previous oxidation studies that thin films can easily be formed on the surface of ferrous alloys; the oxidation rate depends, among other things, on the constitution of the alloy. For example, ferrite oxidizes readily, while the iron carbides resist oxidation. This means that when a specimen of iron containing a carbide is oxidized slightly, there will be a thin film formed over the surface of the sample in which there will be holes wherever a carbide

particle is located. This film may be used now in two ways: (1) Formvar replicas may be made of this "irregular" surface, shadow cast and examined, or (2) the film may be stripped off the specimen and examined directly.

Figures 1c and 1d are micrographs of Formvar replicas of such oxide films formed, in 2 minutes at 500 deg C in air, on samples I and II, respectively. Figure 1c shows a smooth film over the surface of the uncarburized sample. Figure 1d shows the surface of the carburized sample to be roughened during oxidation by the presence of the precipitate particles. The size of the depressions, as depicted in Fig. 1d, is not the true size of the precipitate, since the thickness of the oxide film formed and the subsequent shadow casting have, in this instance, exaggerated the particles size. It is not apparent from these pictures that this is necessarily so; further microscopic examination described in the next paragraph does show it to be the case, however.

Direct observation of the oxide film requires that it be stripped off the metal. This is done by scribing the film in 2-mm squares and placing the metal in a stripping solution. After some 15 seconds the film floats free and is caught on a 200-mesh screen. After 2 rinses the film is dried and used directly in the microscope. The film, as stripped from specimen II, had a fairly smooth substrate with many round holes scattered about in the various grains. These holes are evidently due to the precipitate particles (which have resisted oxidation). The size of these holes should give the true size of the precipitate; such measurements give results in agreement with those made by the Formvar replicas of the electropolished sample.

IV. EXAMINATION OF THE NITROGEN PRECIPITATES

The examination of the nitrogen precipitates was again made mainly by use of Formvar replicas of electropolished samples. The electropolishing was again done using the same conditions as for the samples containing carbon precipitate. The N precipitate was removed from the metal surface in the majority of cases by the electropolish and subsequent rinsing and cleaning operations prior to the making of the replica.

Specimen III, the stable phase Fe_4N, is seen in Fig. 2a. Examination of the replica before and after shadow casting determined that the nitride plate is removed from the metal by the polishing. This replica from the as-polished surface shows that the action of the electropolish develops a structure in the nitride plate. If a sample is then etched chemically, the region in and around a nitride plate is changed to that shown in Fig. 2b. The structure observed in the nitride plate is apparently removed by the action of the etchant. An oxidizing treatment similar to that given samples of specimen II resulted in a structure shown in Fig. 2c. The specimens were not etched prior to oxidation, and the structure confirms that shown in Fig. 2a for the as-polished structure. It indicates that there is a spine in the middle of the nitride plate. A further picture of these nitride plates, which shows even more structure, is Fig. 2d. This picture was made of a plate of Fe_4N which had formed by chance in specimen IV.

The examination of the intermediate N phase proved to be more difficult than for the Fe_4N phase. The electropolish had removed the precipitated phase as before, but the intermediate phase was observed in a variety of shapes. Many regions have the oriented structure as shown in Fig. 3a; however, some regions contain precipitates which have not this regularity (Fig. 3b). The plate-like structure shown in Fig. 3a is the structure predicted for this precipitate by one of the authors.[2] No explanation can be offered at present for the structure shown in Fig. 3b.

Symposium on "Instrumentation for Iron and Steel Industry"

"Instrumentation for the Iron and Steel Industry" will be the subject of a two day conference sponsored by the Carnegie Institute of Technology and the Instrument Society of America to take place at the Roosevelt Hotel, Pittsburgh, Pennsylvania, on March 28–29.

The purpose of this Symposium, which is the first of its type for the steel industry, is to encourage a wider appreciation for and knowledge of instruments in this industry.

The 1951 Symposium on "Instrumentation for the Iron and Steel Industry" is directed toward the man from industry to reveal to him how he can use instruments to improve production, which is so critical in the present emergency. Co-operating societies are American Welding Society, American Society for Metals, American Society for Testing Materials, Association of Iron and Steel Engineers, American Institute of Electrical Engineers, American Society of Mechanical Engineers, Engineers' Society of Western Pennsylvania, and American Institute of Mining and Metallurgical Engineers.

Theory of D_o for Atomic Diffusion in Metals*

CLARENCE ZENER
Institute for the Study of Metals, University of Chicago, Chicago, Illinois
(Received August 18, 1950)

The theory of the entropy of activation ΔS previously given for interstitial diffusion in metals is herein extended to self- and to chemical diffusion. It is concluded that this entropy of activation will always be positive; and it is shown that in the case of self-diffusion and of chemical diffusion at very low concentrations, the better the empirical ΔS the better does it correlate with the theoretical positive ΔS. The large negative apparent ΔS frequently found for chemical diffusion at higher concentrations must be attributed either to inaccuracies in experiments or to the presence of short circuiting diffusion paths. It is predicted that the presently reported negative values of ΔS will become positive when very small concentrations are used in large grain size specimens.

I. INTRODUCTION

THE empirical atomic diffusion coefficient may always be expressed as

$$D(T)=D_o e^{-H/kT}, \tag{1}$$

where D_o and the "heat of activation" H are independent of temperature.

All attempts at obtaining a theoretical expression for D_o start with the following fundamental equation[1] for the three-dimensional random walk problem:

$$D=(1/6)\sum_i \Gamma_i \delta l_i^2. \tag{2}$$

Here Γ_i is the rate at which the diffusing element makes a jump of type i, and δl_i is the corresponding jump distance. The simplicity of the diffusion problem in metals, as contrasted with the diffusion problem in liquids, lies in the fact that in the former all the Γ_i may frequently be taken as equal. Such is the case for self-diffusion and for chemical diffusion at low concentrations. Here we have

$$D=(1/6)\Gamma\sum_i \delta l_i^2. \tag{3}$$

The point of divergence of the various theories for D_o lies in the theoretical expression for the jump rate Γ. In their theoretical study of the diffusion of interstitial solute atoms, Wert and the author[2,3] have found it most convenient to express Γ in the following form

$$\Gamma=\nu e^{-\Delta G/kT}, \tag{4}$$

a form which may be derived in a straightforward manner from statistical mechanics provided ν and ΔG are defined in an appropriate manner. One defines ν as the frequency of vibration of the interstitial atom in an equilibrium position. In case the equilibrium position does not have cubic symmetry, as in b.c.c. (body-centered cubic) lattices, ν refers to the vibration along the direction leading across the saddle configuration. When we come to defining ΔG, we visualize the diffusing interstitial atom confined to a plane normal to the direction leading across the saddle configuration. The isothermal work done in slowly moving this plane from a position passing through an equilibrium position to a position passing through the saddle point, we define as ΔG.

In II we find that Eq. (4) is also applicable to self-diffusion as well as to chemical diffusion in substitutional solutions at low concentrations, ν still denoting a frequency of vibration and ΔG still denoting the work done in a well-defined thought experiment. Numerous other expressions have been proposed for Γ. These are reviewed by Barrer.[4] The advantage of Eq. (4) lies in the fact that ΔG is a well-defined work required to produce a given distortion within the lattice. This clear interpretation of ΔG, in contrast with a definition based upon the ratio of two partition functions, allows us not only to apply rigorous thermodynamic operations upon ΔG, but also enables us to make certain predictions based upon previous experience. When we separate ΔG into an enthalpy and an entropy term

$$\Delta G=\Delta H-T\Delta S, \tag{5}$$

the thermodynamical equation

$$\Delta H=d(\Delta G/T)/d(1/T)$$

ensures us that ΔH is identical to the "heat of activation" H defined by

$$H/k=-d\ln D/d(1/T). \tag{6}$$

Equation (4) may thus be rewritten as

$$\Gamma=\nu e^{\Delta S/k} e^{-H/kT}. \tag{7}$$

A second thermodynamical equation may be used to express the "entropy of activation" ΔS as the negative temperature derivative of ΔG,

$$\Delta S=-d\Delta G/dT. \tag{8}$$

Our experience with the behavior of lattices is now

* This research was supported in part by the ONR (Contract No. N-6ori-20-IV, NR 019 302).

[1] Riemann-Webers, *Differentialgleichungen der Physik* (Braunschweig, 1926), seventh edition, Vol. 2, p. 595.

[2] C. Wert and C. Zener, Phys. Rev. 76, 1169 (1949).

[3] C. Wert, Phys. Rev. 79, 601 (1950).

[4] R. M. Barrer, *Diffusion in and Through Solids* (Cambridge University Press, London, 1941), pp. 291–305.

useful in estimating the value of the right member of Eq. (8).

In previous theoretical treatments of D_o no attempt has been made to estimate ΔS. It has been assumed[5] that in nature both positive and negative values occur, the former corresponding to a loosening of the lattice in the environs of the saddle configuration, the latter to a stiffening. The wide range of the empirical ΔS's, extending from large negative to large positive values, was thus satisfactorily explained. Our Eq. (8) for ΔS allows no such complacent acceptance of this wide range of empirical ΔS values. While it is not impossible that the work ΔG increases with increasing temperature, it is highly improbable. All examples of lattice distortion which have so far been encountered have a negative temperature coefficient for the work done. This negative temperature coefficient of work may be traced ultimately to the negative temperature coefficient of the elastic moduli. Thus, while a negative ΔS cannot be absolutely outruled on theoretical grounds, a negative ΔS is highly improbable. It was the recognition of this essential positive character of ΔS that cast doubt upon the accuracy of those experiments which gave negative ΔS's for interstitial diffusion, and hence which furnished the impetus for the recent precise experiments of Wert.[2,3,6]

Upon combining Eqs. (3) and (7) we obtain

$$D_o=\gamma a^2\nu e^{\Delta S/k}, \tag{9}$$

where a is the lattice parameter, and the numerical coefficient γ is defined by

$$(1/6)\textstyle\sum_i \delta l_i^2=\gamma a^2. \tag{10}$$

This coefficient is completely determined by the geometry of the atomic jumps. Thus for interstitial diffusion in b.c.c. and in f.c.c. (face-centered cubic) lattices, γ is (1/6) and 1, respectively.[2] We find in II that γ is likewise unity in both b.c.c. and f.c.c. lattices for self- and low concentration chemical diffusion, provided such diffusion occurs by the vacancy mechanism.

If all the work ΔG went into straining the lattice, we would anticipate that

$$d(\Delta G/\Delta G_o)/dT=d(\mu/\mu_o)/dT, \tag{11}$$

where μ refers to an appropriate elastic modulus, and the suffix o refers to the absolute zero temperature, or perhaps more correctly, to the extrapolation of the low temperature values to the absolute zero temperature. Since not all of ΔG goes into straining the lattice, we anticipate that the left member of (11) may be somewhat smaller than the right member. We thereby obtain

$$\Delta S=\lambda H d(\mu/\mu_o)/dT, \tag{12}$$

where λ is a numerical coefficient less than, but of the order of magnitude of, unity. In passing from Eqs. (8) and (11) to (12), we have assumed H to be independent of temperature and hence equal to ΔG_o. The current paper by Wert[6] demonstrates that for interstitial solutions, the more accurate the experiments, the more closely does the empirical ΔS satisfy Eq. (12) with λ equal to unity.

TABLE I. Values of $\beta=d(\mu/\mu_0)/d(T/T_m)$.*

Metal	β	Metal	β
Ti	1.1	W	0.35
Th	0.9	Au	0.31
Pb	0.5	Mg	0.31
Ag	0.45	Zn	0.31
α-Fe	0.43 (0.44)[a]	Cd	0.27
Ta	0.40 (0.56)[b]	Ca	0.25
Ba	0.39	La	0.25
Mo	0.36	Pl	0.25
Al	0.35 (0.43)[c]	Be	0.22
Cu	0.35 (0.46)[d]	Pd	0.18

* Values in parentheses are for shear moduli taken from work of Kê; all others are for Young's modulus taken from work of W. Köster, Z. Metallkunde **39**, 1 (1948).
[a] See reference 8. [b] See reference 9. [c] See reference 7. [d] See reference 10.

For ease of computation, the dimensionless quantity

$$\beta=-d(\mu/\mu_o)/d(T/T_m) \tag{13}$$

is given in Table I for various metals, where T_m is the melting temperature. Most of the empirical data on the temperature coefficient of μ have been obtained using polycrystalline specimens. In his extensive measurements Kê[7–10] has demonstrated that the accelerated rate of decrease of μ at elevated temperatures arises from relaxation of shear stress across grain boundaries, that in the absence of such relaxation μ decreases linearly with a rise in temperature. Table I has therefore been constructed from the data in the low temperature range before grain boundary relaxation becomes appreciable. Equation (12) now becomes

$$\Delta S=\lambda\beta H/T_m. \tag{14}$$

This equation is formally similar to the correlation

$$\Delta S=H/T_m-\text{const} \tag{15}$$

recently proposed by Dienes.[11] However, whereas in the Dienes correlation negative ΔS's are allowed, according to the present paper, negative empirical ΔS's either reflect an inaccuracy in the experiments or imply short circuiting paths for diffusion. Possible types of short circuiting paths are discussed in III.

II. SELF-DIFFUSION AND LOW CONCENTRATION CHEMICAL DIFFUSION

In this section we shall consider the diffusing atoms, radioactive isotopes, or chemical solute atoms to be in

[5] A. Stearn and H. Eyring, J. Phys. Chem. **44**, 955 (1940).
[6] C. Wert, J. Appl. Phys. **21**, 1196 (1950).
[7] T. S. Kê, Phys. Rev. **71**, 533 (1947).
[8] T. S. Kê, Trans. Am. Inst. Mining Met. Eng. **176**, 448 (1948).
[9] T. S. Kê, Phys. Rev. **74**, 16 (1948).
[10] T. S. Kê, J. Appl. Phys. **20**, 1226 (1949).
[11] G. J. Dienes, J. Appl. Phys. **21**, 1189 (1950).

such low concentration that they suffer no appreciable mutual interference in their diffusion.

We shall first investigate the vacancy method of migration. Then in the case of self-diffusion the rate Γ at which a radioactive isotope atom makes a jump into a specific nearest neighbor site may be written as

$$\Gamma = \Gamma' c,$$

where c is the probability that the site be vacant, i.e., the atomic concentration of holes, and Γ' is the rate of jumping when a hole is present. Now

$$c = e^{-\Delta G'/kT},$$

where $\Delta G'$ is the isothermal work necessary to form a lattice vacancy in a particular lattice site. When we now denote by $\Delta G''$ the work necessary to move our constraining plane from the potential minimum to the saddle point, and set

$$\Delta G = \Delta G' + \Delta G'',$$

we obtain Eq. (4). We have already seen that $\Delta S''$ is positive. It is readily seen that $\Delta S'$ is likewise positive. When a vacancy is formed at a particular lattice site, the neighboring atoms acquire more freedom of movement, and must thus necessarily increase their entropy when the vacancy is formed under isothermal conditions. In a f.c.c. lattice there are 12 possible jump directions, and each jump has a length of $2^{-\frac{1}{2}}a$. From Eq. (10) we thereby find that γ is unity. In a b.c.c. lattice there are 8 possible jump directions, and each jump has a length of $(1/2)3^{\frac{1}{2}}a$. We thereby obtain that in a b.c.c. lattice also, γ is unity for self-diffusion by the vacancy mechanism.

In the case of chemical diffusion a complication arises when the work $\Delta G'$ necessary to form a vacancy in a particular lattice site adjacent to a solute atom is appreciably smaller than the work necessary to form a vacancy in a particular lattice site away from all solute atoms. In such a case, once a vacancy becomes a nearest neighbor of a solute atom, it will change places with this solute atom, and with lattice atoms which are nearest neighbors of the solute atom and itself, many times before it again wanders off into the surrounding lattice. Such a possibility was first suggested by Johnson.[12] In analyzing this situation, we set

$$D = fD_p,$$

where f is that fraction of solute atoms which are paired with vacancies, and D_p is the diffusion coefficient of the center of gravity of a solute atom-vacancy molecule. Now

$$f = \sigma e^{-\Delta G'/kT},$$

where σ is the number of nearest neighboring lattice sites, and $\Delta G'$ is the work necessary to form a vacancy on a particular site adjacent to a solute atom. In the case of a f.c.c. lattice such a vacancy can jump to 4 nearest neighboring sites which are also nearest neighboring sites to the solute atom. Each such jump of length δl is associated with a jump of the center of the solute atom-vacancy molecule of $\delta l/2$. We thereby again arrive at Eq. (9) for D_o with γ equal to unity. In the case of a b.c.c. lattice the solute atom and vacancy have no nearest neighboring lattice site in common. It does not therefore appear that solute atom-vacancy molecules will diffuse in b.c.c. lattices. In such lattices the diffusion rate of the solute atoms will be dependent upon the rate of migration of the vacancies through the lattice. The rates of chemical diffusion and of self-diffusion will then become essentially identical. No data are available at present with which to test this conclusion.

A second likely mechanism of diffusion is by the rotation of a ring of atoms, preferably a 4-atom ring in f.c.c. and in b.c.c. lattices.[13] The rate at which a particular atom jumps in a particular direction to a neighboring lattice site is then $n\Gamma''$, where n is the number of distinct rings which will effect this jump, and Γ'' is the rate at which each ring rotates $\pi/2$ in the appropriate direction. In a f.c.c. lattice, n is 2. Thus, an atom initially at $(1/2, 0, 1/2)a$ can jump to $(0, 1/2, 1/2)a$ by the rotation of two distinct rings lying in the xy plane passing through these two lattice sites. In a b.c.c. lattice, n is 6. Thus, an atom initially at $(0, 0, 0)a$ can jump to $(1/2, 1/2, 1/2)a$ by a rotation of two distinct rings in each of the three planes $(0, 1, \bar{1})$, $(1, \bar{1}, 0)$, and $(\bar{1}, 0, 1)$. The analysis of the statistical mechanics of ring rotation follows, in every respect, that used for the diffusion of interstitial atoms, save that now we constrain that normal coordinate of the ring corresponding to a net rotation rather than that coordinate of the interstitial atom which passes over the saddle point. We thus obtain Eq. (9) for D_o, γ being 2 and 6 for f.c.c. and b.c.c. lattices, respectively. If we thus define the experimental ΔS as

$$(\Delta S/k)_{\text{exp}} = \ln(D_o/a^2\nu), \tag{16}$$

$(\Delta S/k)_{\text{exp}}$ will be greater by the amount ln2 or ln6 for diffusion by ring rotation than by vacancy migration. Therefore, $(\Delta S)_{\text{exp}}$ should be positive irrespective of the precise mechanism of diffusion.

All the experimental data on self-diffusion, and on chemical diffusion at very low concentrations, are presented in Fig. 1. We see that in every case $(\Delta S)_{\text{exp}}$ is positive. We likewise see that in every case of either self-diffusion or of low concentration chemical diffusion in f.c.c. lattices, the most recent data are consistent with Eq. (14) for ΔS with the coefficient λ equal to 0.55. The most recent data for self-diffusion in b.c.c. lattices are consistent with this equation with λ equal to unity. The data[14] for self-diffusion in the high temperature

[12] R. P. Johnson, Phys. Rev. **56**, 814 (1939).

[13] C. Zener, Acta Cryst. **3**, 346 (1950).

[14] Buffington, Bakalar, and Cohen (private communication).

phase (f.c.c.) of iron have not been plotted because of the lack of data on the temperature coefficient of μ in the absence of grain boundary relaxation. $(\Delta S/k)_{exp}$ for this system is 3.9 and would be consistent with λ equal to 0.55 if β for this system had the value of 0.37.

The difference in the manner in which the data for f.c.c. and for b.c.c. lattices are correlated in Fig. 1 suggests a difference in mechanism of diffusion. The higher values of ΔS for the b.c.c. lattice suggests, in fact, the ring mechanism for that lattice, the vacancy mechanism for the f.c.c. lattice. The recent experiments of Kuczynski[15] appear to demand, in fact, that at least in the f.c.c. metal silver, self-diffusion occurs by the vacancy mechanism. The rate of sintering which he computes according to the vacancy mechanism agrees remarkably well with that which he observes.

III. SHORT CIRCUIT DIFFUSION PATHS

In the previous section we have considered all lattice sites as equivalent. When the matrix is inhomogeneous, the possibility exists that there are certain paths along which diffusion will have a lower heat of activation than through the bulk of the matrix. Since at any one time only a fraction of the solute atoms will be in these paths of rapid diffusion, the associated D_o will be less than that corresponding to homogeneous diffusion. If the major part of the diffusion occurs along these preferred paths, and we persist in defining the experimental ΔS by Eq. (16), then this ΔS will be smaller than the theoretical value, and may even be negative. We are thereby led to interpret all cases of negative $(\Delta S)_{exp}$ in terms of inhomogeneities which furnish short circuiting diffusion paths.

These inhomogeneities may be inherent in the structure. Thus, grain boundaries have been found to furnish paths for rapid diffusion in certain cases, notably in the diffusion of thorium in tungsten.[16,17] Structural inhomogeneities introduced by plastic deformation, which may be interpreted as dislocations, have also been found to accelerate diffusion.[18]

The inhomogeneity may also be chemical in nature. Thus, let us suppose that in a solid solution of zinc in copper diffusion occurs by the vacancy mechanism, and that the vacancies change places much more readily with zinc atoms than with copper atoms. If the concentration of zinc is not too small, the vacancies will then diffuse essentially through a continuous network of zinc atoms. Such a network will not exist at very low concentrations.

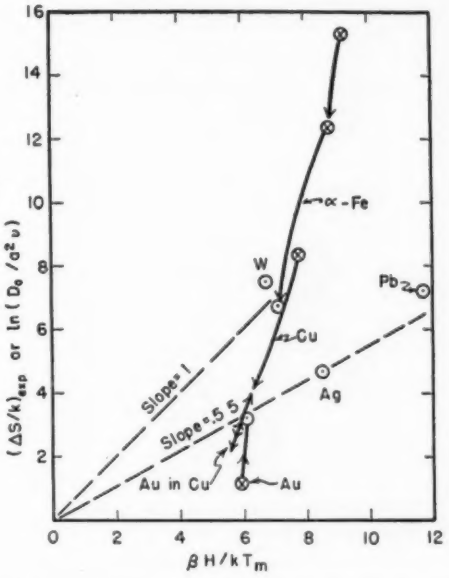


FIG. 1. Comparison of experimental values of $\Delta S/k$ with $\beta(H/kT_m)$ for self-diffusion of cubic metals and for chemical diffusion in very dilute solutions as determined by radioactive tracer technique. Successive experimental values on the same system are taken from the following references. Alpha-iron: C. E. Birchenall and R. F. Mehl, J. Appl. Phys. **19**, 217 (1948), C. E. Birchenall and R. F. Mehl, Trans. Am. Inst. Mining Met. Eng. **188**, 144 (1950); Buffington, Bakalar, and Cohen, see reference 14. Copper: Steigman, Shockley, and Nix, Phys. Rev. **56**, 13 (1939), M. S. Maier and H. R. Nelson, Trans. Am. Inst. Mining Met. Eng. **147**, 39 (1942). Gold: H. A. C. McKay, Trans. Faraday Soc. **34**, 845 (1938), surface technique, A. M. Sagrubskij, Physik Z. Sowjetunion **12**, 118 (1937), slicing technique. Au in Cu: A. B. Martin and F. Asaro, Phys. Rev. **80**, 123 (1950).

The above postulated association of negative values of $(\Delta S)_{exp}$ with inhomogeneities leads to definite predictions. If all inhomogeneities which may give rise to short circuiting diffusion paths are removed, then the $(\Delta S)_{exp}$ will be raised to the theoretical positive range. We thus anticipate that all reported negative values of $(\Delta S)_{exp}$ for chemical diffusion will become positive when very small concentrations are used in specimens having sufficiently large grain size so that grain boundaries are unimportant, and in specimens having had sufficiently long anneals at a high temperature so that diffusion along dislocations is inappreciable.

[15] G. C. Kuczynski, J. Appl. Phys. **21**, 632 (1950).
[16] Fonda, Young, and Walker, Physics **4**, 1 (1933).
[17] I. Langmuir, J. Franklin Inst. **217**, 543 (1934).
[18] C. Matanno, Japan, J. Phys. **9**, 41 (1934).

On the Scattering of Plane Waves by Soft Obstacles. I. Spherical Obstacles*

ROBERT W. HART†
University of Pittsburgh, Pittsburgh, Pennsylvania, and The Catholic University of America, Washington, D. C.

AND

ELLIOTT W. MONTROLL‡
University of Pittsburgh, Pittsburgh, Pennsylvania

(Received August 9, 1950)

An approximate theory is developed for the scattering of plane waves by spherical obstacles with the property $m=\lambda_0/\lambda_1<1.5$, where λ_1 and λ_0 are, respectively, the wavelengths of a plane wave in the scattering material and in the surrounding medium. In the exact theory the scattered field is found to be an infinite series of legendre polynomials whose coefficients are complicated combinations of Bessel functions. By making an approximation in the nth coefficient that is valid when either $|m-1|\ll 1$ or $2\pi a/\lambda_0\gg n$ (a=sphere diameter) it is possible to sum the series and obtain analytical closed form expressions for the total scattering cross section of the sphere and the intensity of the scattered wave as a function of angle. These approximate expressions are compared in the optical case with the results obtained numerically by the Bureau of Standards Computing Laboratory. As an example of their accuracy and range, it is found that when $m<1.1$ the relative error in the approximate total scattering cross section is less than 5 percent for scatterers of any size; when $1.1<m<1.5$, the relative error is less than 5 percent if $2\pi a(m-1)/\lambda_0<6$ and less than 25 percent for any size sphere.

I. INTRODUCTION

WE shall qualitatively consider an obstacle to be soft to a plane wave if it subjects the wave to only a small phase shift. A spherical dielectric particle whose refractive index is close to that of its surrounding medium is soft to a light wave. The virus proteins (for example, bushy stunt virus and influenza virus) are soft to visible light. The relative index of scatterer to medium is between 1.2 and 1.3 in these examples. A spherical obstacle is soft to a sound wave if the product ρc (ρ=density and c=velocity of sound) in the sphere does not differ much from that of its surrounding medium.

Measurements on the scattering of light by soft obstacles are usually interpreted through the Rayleigh-Gans theory.[1–4] There, the electric field associated with the light wave at a point in the scatterer is assumed to have the same value that it would have had in the absence of the scatterer. The corresponding theory for the scattering of scalar waves is called the Born[5] approximation. The most striking fault of this theory lies in the way it predicts the total scattering cross section of the obstacle to become a monatonically increasing function of the ratio of one of its characteristic lengths to the wavelength of the scattered light as this ratio increases. In a correct theory, the total scattering cross section oscillates and finally approaches an asymptotic value of the same order of magnitude as the geometrical cross section of the scatterer; that is, the value obtained by geometrical optics: $2\pi(\text{radius})^2$ for a sphere.[6–8] This deviation from the exact results is not surprising, because the Rayleigh-Gans approximation of the internal by the external field is far from the truth in a large scatterer.

The exact solution of the wave equation for the scattering of scalar waves by spheres has been obtained by Rayleigh[9] and for electromagnetic waves by Mie.[10] Unfortunately, these solutions exist as infinite series in legendre polynomials whose coefficients are complicated combinations of Bessel functions. Calculations based on these series become very tedious. Even though the new Bureau of Standards tables of scattering functions[7,11] should prove to be a useful aid in scattering calculations, it is still desirable to obtain simple approximate formulas for these functions.

One of the purposes of this paper is to derive a new closed form approximation for the Rayleigh and Mie series. This approximation yields total scattering cross sections which are in good agreement with the exact results when the relative index of scatterer, m, is in the range 1.0–1.5. The total scattering cross section remains finite for larger particles; and as $m\to 1$ and $2\pi a/\lambda\to\infty$, it correctly approaches twice the geometrical cross section. Unfortunately, our results on polarization by scattering are not in such good agreement with the Mie

* Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of Pittsburgh.

† Present address: Johns Hopkins Applied Physics Laboratory, Silver Spring, Maryland.

‡ Present address: Institute of Fluid Dynamics and Applied Mathematics, University of Maryland, College Park, Maryland.

[1] Lord Rayleigh (J. W. Strutt), Proc. Roy. Soc. (London) **84A**, 25 (1911); **90**, 219 (1914).

[2] R. Gans, Ann. Physik **76**, 29 (1925).

[3] P. Debye, J. Appl. Phys. **15**, 338 (1944).

[4] Zimm, Stein, and Doty, Polymer Bull. **1**, 90 (1945).

[5] M. Born, Z. Physik **37** (1926), and **38**, 803 (1926).

[6] H. C. van de Hulst, Recherches Astronomiques de l'observatoire d'Utrecht **11** (1946).

[7] D. Sinclair, J. Opt. Soc. Am. **37**, 475 (1947).

[8] L. Brillouin, J. Appl. Phys. **20**, 1108 (1949).

[9] Lord Rayleigh, *The Theory of Sound* (Dover Publications, New York, 1945), reprint edition, Vol. 2, p. 272.

[10] G. Mie, Ann. Physik **25**, 37 (1908).

[11] *Tables of Scattering Functions for Spherical Particles* (National Bureau of Standards Applied Mathematics Series **4**, Washington, 1948).

theory. Our theory reduces to a new result of van de Hulst[6] for the limiting scattered intensity as $m \to 1$.

Some further new results in scattering and diffraction by spheres have been obtained by van de Hulst[12] and Ljunggren.[13]

For further references and a rather complete review of recent progress in both the experimental and theoretical aspects of light scattering, the reader is referred to a paper of G. Oster.[14] A detailed exposition of scattering of scalar waves has been given by Mott and Massey.[15]

It will be assumed throughout this paper that there is no absorption of the incident radiation by the obstacle. However, there does not seem to be any reason why our results should not be valid for an absorbing sphere whose complex index of refraction is of absolute value less than 1.5.

II. RAYLEIGH-GANS THEORY

In this section we shall summarize the Rayleigh-Gans theory of scattering by a sphere of uniform density and by one of gaussian density.

Let us consider a polarized plane wave of wavelength λ propagated in an infinite homogeneous isotropic medium in the direction of the unit vector $\mathbf{s}_0$ and polarized in the direction of a unit vector $\mathbf{p}$. The electric field associated with the wave at a point P is given by

$$\mathbf{E}_0(r) = E_0 \mathbf{p} \exp(-i\omega t) \exp(-2\pi i \mathbf{r}\cdot\mathbf{s}_0/\lambda). \quad (1)$$

Here ω is the angular frequency of the wave, E_0 is its scalar amplitude, and $\mathbf{r}$ is the vector distance of P from an origin at which the phase is zero.

Let a homogeneous isotropic dielectric particle small compared with λ, of volume δV and volume polarizability α, be placed at P. Its polarization will be in phase with the external field, its induced dipole moment being

$$\delta\mathbf{u} = \alpha \mathbf{E}_0(r)\delta V. \quad (2)$$

The dipole radiation emitted from the particle is to be identified with the electric field of the scattered light. The phase of the scattered radiation at point Q is

$$-(2\pi/\lambda)(\mathbf{R}\cdot\mathbf{s}_1 + \mathbf{r}\cdot\mathbf{s}_0) = (2\pi r/\lambda)(\mathbf{s}_1 - \mathbf{s}_0) - (2\pi \mathbf{R}'/\lambda)\cdot\mathbf{s}_1,$$

where $\mathbf{s}_1$ is a unit vector in the direction QP. From classical electrodynamics the electric field of the scattered radiation at Q is

$$\mathbf{E} = (4\pi^2 E_0/R\lambda^2)[\mathbf{s}_1\times(\mathbf{s}_1\times\mathbf{p})]\alpha e^{-i\beta} \times \exp[2\pi i \mathbf{r}\cdot(\mathbf{s}_1-\mathbf{s}_0)/\lambda], \quad (3)$$
$$\beta = \omega(t - |R|/c) + 2\pi \mathbf{R}'\cdot\mathbf{s}_1/\lambda.$$

Since the light intensity of an electromagnetic wave is proportional to $|\mathbf{E}|^2$, the ratio of the scattered intensity at Q polarized in a direction $\mathbf{q}$ to the intensity of the original plane wave is

$$I_s/I_0 = \frac{(2\pi/\lambda)^4}{R^2}(\mathbf{p}\cdot\mathbf{q})^2\alpha^2(\delta V)^2. \quad (4)$$

This is the famous Rayleigh law of scattering of a plane wave by a small particle.

Now, let n small dielectric particles be located at $\mathbf{r}_1, \cdots, \mathbf{r}_n$ (a single large particle can be envisioned as a large number of adjoining smaller ones), where $\max|\mathbf{r}_i - \mathbf{r}_j| \ll R$. Also, let $\mathbf{E}(\mathbf{r}_j)$ be the electric field at $\mathbf{r}_j$ and $\alpha(\mathbf{r}_j)$ the polarizability of jth particle. The jth induced dipole moment is $\delta\mathbf{u}_j = \alpha(\mathbf{r}_j)\mathbf{E}(\mathbf{r}_j)\delta V_j$. $\mathbf{E}(\mathbf{r}_j)$ includes the effect of scattered waves as well as the original wave. Hence, in general $\mathbf{E}(\mathbf{r}_j) \neq \mathbf{E}_0(\mathbf{r}_j)$. However, as $\alpha \to 0$ $\mathbf{E}(\mathbf{r}_j) = \mathbf{E}_0(\mathbf{r}_j)[1+O(\alpha)]$. Hence, in a very soft scatterer where α is small (that is, where the relative index m of scatterer to medium is close to unity) one can use the Rayleigh-Gans approximation $\mathbf{E}(\mathbf{r}_j) = \mathbf{E}_0(\mathbf{r}_j)$ to obtain formulas correct to first power in α or $(m-1)$.

This approximation leads to the following field at the point Q:

$$\mathbf{E} = (4\pi^2 E_0/R\lambda^2)[\mathbf{s}_1\times(\mathbf{s}_1\times\mathbf{p})]e^{-i\beta}\sum_j \alpha(\mathbf{r}_j) \times \exp[2\pi i \mathbf{r}_j\cdot(\mathbf{s}_1-\mathbf{s}_0)/\lambda]\delta V_j. \quad (5a)$$

When the points $\{\mathbf{r}_j\}$ represent the locations of the infinitesimal volume elements of a large chunk of dielectric, the summation in Eq. (5a) can be replaced by an integration. The scattered intensity at Q is then

$$I_s/I_0 = R^{-2}(2\pi/\lambda)^4(\mathbf{p}\cdot\mathbf{q})^2 \times \left|\int_V \alpha(\mathbf{r}) \exp[2\pi i \mathbf{r}\cdot(\mathbf{s}_1-\mathbf{s}_0)/\lambda]dv\right|^2. \quad (5b)$$

An important error which appears in Eq. (5b) is the absence of a component of scattered light polarized perpendicular to $\mathbf{p}$ (i.e., if $\mathbf{q}$ is perpendicular to $\mathbf{p}$, then $I_s = 0$). In an exact theory this component is small but finite.

The integral in Eq. (5b) is easily obtained for a spherical obstacle of uniform polarizability α and radius a. We introduce a new coordinate system whose origin is at the center of the sphere and whose polar axis is in the direction of $(\mathbf{s}_1-\mathbf{s}_0)$. Let γ be the polar angle formed by $(\mathbf{s}_1-\mathbf{s}_0)$ and the vector connecting the center of the sphere with the variable point $\mathbf{r}$ that ranges over its interior. If $\mathbf{r}'$ is the vector which connects $\mathbf{r}$ with the origin of the new coordinate system, and r_0 is the vector which connects the origin of the new coordinate system with that of the old one, the required integral is then

$$I = 2\pi\alpha\int_0^a\int_0^\pi \exp[2\pi i \mathbf{r}\cdot(\mathbf{s}_1-\mathbf{s}_0)/\lambda]r'^2 \sin\gamma dr' d\gamma;$$

but $\mathbf{r}\cdot(\mathbf{s}_1-\mathbf{s}_0) = \mathbf{r}_0\cdot(\mathbf{s}_1-\mathbf{s}_0) + \mathbf{r}'\cdot(\mathbf{s}_1-\mathbf{s}_0) = \mathbf{r}_0\cdot(\mathbf{s}_1-\mathbf{s}_0) + 2r'\cos\gamma\sin\tfrac{1}{2}\theta$, where θ is the angle between $\mathbf{s}_0$ and $\mathbf{s}_1$.

[12] H. C. van de Hulst, Recherches Astronomiques de l'observatoire d'Utrect **11**, Part 2 (1949).
[13] T. Ljunggren, Ark. Mat. Astron. Fysik **36A**, No. 14 (1948); Arkiv Fysik **1**, 1 (1949).
[14] G. Oster, Chem. Revs. **43**, 319 (1948).
[15] Mott and Massey, *Theory of Atomic Collisions* (Oxford University Press, London, 1933).

Hence, if

$$\epsilon=\exp[2\pi i\mathbf{r}_0\cdot(\mathbf{s}_1-\mathbf{s}_0)],$$

then

$$I=2\pi\alpha\epsilon\int_0^a\int_0^\pi r'^2 \times\exp[(4\pi ir'/\lambda)\sin\tfrac{1}{2}\theta\cos\gamma]\sin\gamma d\gamma dr'$$
$$=4\pi\alpha\epsilon a^3(\tfrac{1}{2}\pi)^{\frac{1}{2}}J_{\frac{3}{2}}[(4\pi a/\lambda)\sin\tfrac{1}{2}\theta]/[(4\pi a/\lambda)\sin\tfrac{1}{2}\theta]^{\frac{3}{2}}, \quad (6)$$

where $J_{\frac{3}{2}}(z)$ is the Bessel function of order $\frac{3}{2}$. If we introduce the classical relation between polarizability α and index of refraction m, $(m^2-1)/4\pi=\alpha$, and substitute Eq. (6) into Eq. (5b), we obtain the Rayleigh-Gans expression:

$$I_s/I_0=(\pi a^2/2R^2)x^4(\mathbf{p}\cdot\mathbf{q})^2(m^2-1)^2\{J_{\frac{3}{2}}(\rho)/\rho^{\frac{3}{2}}\}^2; \quad (7a)$$

$$x=2\pi a/\lambda; \quad \rho=2x\sin\tfrac{1}{2}\theta. \quad (7b)$$

Scattering by a gaussianly dispersed sphere can be described to the Rayleigh-Gans approximation rather easily. Many polymers are approximated by a gaussian distribution of monomer elements. Let us consider the case,

$$\alpha(r')=\alpha_0\exp[-(r'/a)^2]. \quad (8a)$$

The integral in Eq. (5b) is

$$I=2\pi\alpha_0\epsilon\int_0^\infty\int_0^\pi r'^2\exp[-(r'/a)^2] \times\exp[(4\pi ir'/\lambda)\sin\tfrac{1}{2}\theta\cos\gamma]\sin\gamma dr'd\gamma$$
$$=(\lambda\alpha_0\epsilon/\sin\tfrac{1}{2}\theta)\int_0^\infty r'\exp[-(r'/a)^2] \times\sin[(4\pi r'/\lambda)\sin\tfrac{1}{2}\theta]dr'$$
$$=\alpha_0\pi^{\frac{1}{2}}a^3\epsilon\exp\{-[(\pi a^2/\lambda^2)\sin^2\tfrac{1}{2}\theta]\}.$$

Hence,

$$\frac{I_s}{I_0}=\left(\frac{a}{R}\right)^2\left(\frac{2\pi a}{\lambda}\right)^4(\mathbf{p}\cdot\mathbf{q})^2\alpha_0^2\pi^3 \times\exp\left[-\left(\frac{2\pi a^2}{\lambda^2}\sin^2\tfrac{1}{2}\theta\right)\right]. \quad (8b)$$

If the initial plane wave is unpolarized, it can be represented as the superposition of two incoherent waves whose polarizations are perpendicular to each other and to the direction of wave propagation. The total scattered intensity for both directions of polarization can be shown to be $\frac{1}{2}(1+\cos^2\theta)(I_s/I_0)_{\mathbf{p}\cdot\mathbf{q}=1}$. The total scattering cross section for spherically symmetrical particles is

$$\sigma_s=2\pi R^2\int_0^\pi\tfrac{1}{2}(1+\cos^2\theta)(I_s/I_0)_{\mathbf{p}\cdot\mathbf{q}=1}\sin\theta d\theta. \quad (9)$$

This quantity can be computed for both the homogeneous and dispersed spheres.

(A) Homogeneous Sphere

The integration over θ can be replaced by one over ρ. Then,

$$\sigma_s=\tfrac{1}{2}\pi^2a^2x^2(m^2-1)^2\int_0^{2x}\rho^{-2}J_{\frac{3}{2}}^2(\rho) \times[2-(\rho/x)^2+(\rho^4/4x^4)]d\rho.$$

The required integrals have been obtained by Lord Rayleigh so that

$$\sigma_s=\tfrac{1}{2}\pi^2a^2x^2(m^2-1)^2[2I_1-(I_2/x^2)+I_3/4x^4], \quad (10a)$$

where

$$I_1(2x)=\int_0^{2x}\rho^{-2}J_{\frac{3}{2}}^2(\rho)d\rho =(1/2\pi)-(1/8x)\{J_{\frac{3}{2}}^2(2x)+J_{\frac{1}{2}}^2(2x)\}, \quad (10b)$$

$$I_2(2x)=\int_0^{2x}J_{\frac{3}{2}}^2(\rho)d\rho =(1/\pi)\{(\sin4x/2x)-(\sin^22x/4x^2)-1+\varphi(4x)\}, \quad (10c)$$

$$I_3(2x)=\int_0^{2x}\rho^2J_{\frac{3}{2}}^2(\rho)d\rho =(1/\pi)\{2x^2+x\sin4x-(5/2)\sin^22x+\varphi(4x)\}, \quad (10d)$$

$$\varphi(z)=\int_0^z(1-\cos y)dy/y. \quad (10e)$$

By this approximation $(\sigma_s/\pi a^2)$ is a monatonically increasing function of x rather than a bounded oscillating one as derived by a more exact theory (Fig. 2).

(B) Gaussianly Dispersed Sphere

In this case it is desirable to integrate over $y=-\cos\theta$. Then,

$$\sigma_s=a^2(2\pi a/\lambda)^4\alpha_0^2\pi^4 \times\exp(-\pi a^2/\lambda^2)\int_{-1}^1\exp(-\pi a^2y/\lambda^2)(1+y^2)dy$$
$$=16a^2\pi^5\alpha_0^2(2\pi a/\lambda)^2 \times\exp(-\pi a^2/\lambda^2)\{(1+\lambda^4/\pi^2a^4)\sinh(\pi a^2/\lambda^2) -(\lambda^2/\pi a^2)\cosh(\pi a^2/\lambda^2)\}. \quad (11)$$

We shall now derive more exact expressions for the scattering functions than those of Rayleigh and Gans. First, we shall deal with scattering of scalar waves. Then the more complicated electromagnetic waves will be investigated.

III. SCATTERING OF PLANE SCALAR WAVES

Let a plane scalar wave of wavelength λ_0 be scattered by an isotropic non-absorbing sphere composed of a material in which its wavelength is λ_1. If the sphere is of radius a and has its center at the origin, the wave functions of the incident, scattered, and transmitted waves all satisfy the wave equations

$$\nabla^2\psi+k_1^2\psi=0 \quad \text{if} \quad r<a \quad \text{and} \quad k_1=2\pi/\lambda_1, \tag{12a}$$

$$\nabla^2\psi+k_0^2\psi=0 \quad \text{if} \quad r>a \quad \text{and} \quad k_0=2\pi/\lambda_0. \tag{12b}$$

At the boundary of the scatterer, $r=a$, the quantities $(r\psi)_{r=a}$ and $[\partial(\psi r)/\partial r]_{r=a}$ must be continuous. We shall find approximate solutions of the wave equation and scattering cross section for scatterers such that $|k_1-k_0|$ is small, but not necessarily as small as is required in the range of validity of the Rayleigh-Gans-Born approximation.

In the case of particle waves $k^2=2m(E-V)/\hbar^2$ (m=mass of particle, $E-V$=difference between energy of particle and constant potential energy of force acting on particle, $\hbar$=(Planck's constant)$/2\pi$, and ψ is the Schroedinger wave function). In the acoustical case ψ can be identified with the time-independent factor of the pressure associated with the wave on the hypothesis that $p=\psi\exp(-2\pi i\nu t)$. The sound intensity given by $|\psi|^2/2\rho c$ (ρ=density of medium and c=sound velocity). Our boundary conditions correspond to scattering of sound waves by an obstacle whose density is the same as that of the surrounding medium.

The solutions of the wave equation in spherical coordinates are

$$\psi_n^{(m)}=z_n(kr)P_n(\cos\theta)e^{\pm im\varphi}; \quad n=0, 1, \cdots; m=0, 1, \cdots, n, \tag{12c}$$

where $z_n(kr)$ is the spherical Bessel function defined by

$$z_n(u)=(\pi/2u)^{\frac{1}{2}}Z_{n+\frac{1}{2}}(u).$$

Z represents either J (Bessel) or $H^{(1)}$ (Hankel) function. Since the solutions (12c) form a complete set of orthogonal functions, we can express the incident, scattered, and transmitted waves as linear combinations of these functions.

Let the incident plane wave be propagated with unit amplitude in the direction of the z axis. Then,

$$\psi_i=e^{irk_0\cos\theta}=\sum_{n=0}^{\infty} i^n(2n+1)P_n(\cos\theta)j_n(rk_0). \tag{13a}$$

We expand the ψ for the scattered wave in the series

$$\psi_s=\sum_{n=0}^{\infty} A_{s,n}i^n(2n+1)h_n(rk_0)P_n(\cos\theta). \tag{13b}$$

Inasmuch as ψ_i is independent of φ, so also is ψ_s, and the transmitted wave

$$\psi_t=\sum_{n=0}^{\infty} A_{t,n}i^n(2n+1)j_n(rk_1)P_n(\cos\theta). \tag{13c}$$

The constants $A_{n,s}$ and $A_{t,n}$ are found by applying the boundary conditions given in the text below Eq. (12b). One finds that

$$A_{t,n}j_n(ak_1)=A_{s,n}h_n(ak_0)+j_n(ak_0), \tag{14a}$$

$$A_{t,n}[ak_1j_n(ak_1)]'=A_{s,n}[ak_0h_n(ak_0)]'+[ak_0j_n(ak_0)]'. \tag{14b}$$

By solving these equations for $A_{s,n}$ and employing the relation $[xj_n(x)]'=(n+1)j_n(x)-xj_{n+1}(x)$, one obtains

$$-A_{s,n}=\frac{k_1j_n(ak_0)j_{n+1}(ak_1)-k_0j_n(ak_1)j_{n+1}(ak_0)}{k_1h_n(ak_0)j_{n+1}(ak_1)-k_0j_n(ak_1)h_{n+1}(ak_0)}. \tag{15}$$

If the numerator is rewritten in terms of ordinary Bessel functions of half-integral order, it can be simplified (see Eq. (8), p. 134, Watson's *Bessel Functions*) to

$$\tfrac{1}{2}\pi[(k_1^2-k_0^2)/a^2(k_1k_0)^{\frac{1}{2}}]\int_0^a xJ_{n+\frac{1}{2}}(xk_1)J_{n+\frac{1}{2}}(xk_0)dx. \tag{16}$$

In order to sum the series (13b) in closed analytical form, we shall examine several approximate forms of the denominator of Eq. (15). A rather simple expression can be obtained when k_1a, $k_0a\gg n$. When[16] $x\gg n$,

$$h_n(x)\simeq(-i)^{n+1}e^{ix}/x, \quad j_n(x)\simeq x^{-1}\cos[x-\tfrac{1}{2}(n+1)\pi]. \tag{17}$$

Under the above condition, the denominator of Eq. (15) becomes

$$[i(k_1+k_0)/2k_1k_0a^2]e^{-i(k_1-k_0)a}[1+(-1)^n\kappa e^{2ik_1a}], \tag{18a}$$

and Eq. (15) is approximated by

$$A_{s,n}\simeq i\pi(k_1k_0)^{\frac{1}{2}}(k_1-k_0)e^{i(k_1-k_0)a}[1-(-1)^n\kappa e^{2ik_1a}] \times\int_0^a xJ_{n+\frac{1}{2}}(k_1x)J_{n+\frac{1}{2}}(k_0x)dx/(1-\kappa^2e^{4ik_1a}), \tag{18b}$$

where

$$\kappa=(k_1-k_0)/(k_1+k_0).$$

Let us now seek an approximation of the denominator of Eq. (15) when $k_1\simeq k_0$. Then it becomes[17] ($p=\frac{1}{2}+n$)

$$\tfrac{1}{2}k_0(\pi/ak_0)[H_p^{(1)}(ak_0)J_{p+1}(ak_0)-J_p(ak_0)H_{p+1}^{(1)}(ak_0)]=i/k_0a^2. \tag{18c}$$

However, this is exactly Eq. (18a) in the limit $k_1\to k_0$. Hence, Eqs. (18a) and (18b) are valid under both sets of conditions $|(k_0/k_1)-1|\ll 1$ and k_1a, $k_0a\gg n$. The first of these is the "soft sphere" condition. The second gives us some hope that our resulting expressions may have a wider range of application than those of Rayleigh, Gans, and Born (especially for larger spheres).

At distances sufficiently far from the scatterer so that the asymptotic expression (17) can be used for $h_n(rk_0)$,

[16] J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill Book Company, Inc., New York, 1941), p. 406.

[17] Jahnke-Emde, *Funktionentafeln* (Leipzig, 1933), p. 212.

one can easily approximate the sum (13b) by using Eq. (18b) and the series[18]

$$\sin x\omega/\omega=(\pi x/2\omega)^{\frac{1}{2}}J_{\frac{1}{2}}(x\omega)$$

$$=\frac{\pi}{2(k_1k_0)^{\frac{1}{2}}}\sum_{n=0}^{\infty}(2n+1)J_{n+\frac{1}{2}}(xk_0)\times J_{n+\frac{1}{2}}(xk_1)P_n(\cos\theta), \quad (19a)$$

$$\sin xv/v=(\pi x/2v)^{\frac{1}{2}}J_{\frac{1}{2}}(xv)$$

$$=\frac{\pi}{2(k_1k_0)^{\frac{1}{2}}}\sum_{n=0}^{\infty}(-1)^n(2n+1)J_{n+\frac{1}{2}}(xk_0)\times J_{n+\frac{1}{2}}(xk_1)P_n(\cos\theta), \quad (19b)$$

$$v^2=k_0^2+k_1^2+2k_0k_1\cos\theta,$$
$$\omega^2=k_0^2+k_1^2-2k_0k_1\cos\theta. \quad (19c)$$

Since

$$\int_0^b y^{\frac{1}{2}}J_{\frac{1}{2}}(y)dy=b^{\frac{1}{2}}J_{\frac{1}{2}}(b), \quad (19d)$$

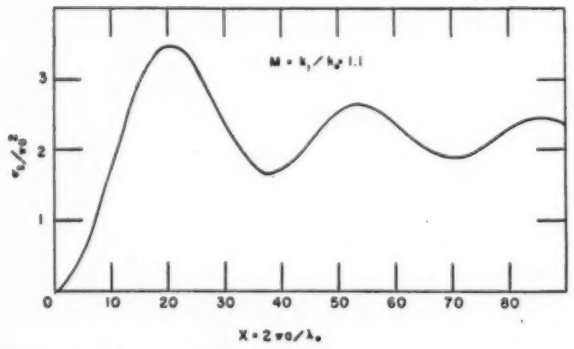


FIG. 1. Ratio of scattering cross section, σ_s, to sphere cross section for a scalar plane wave with $m=k_1/k_0=1.1$ as a function of the sphere size parameter $x=2\pi a/\lambda_0$. Here, $k_j=2\pi/\lambda_j$, λ_0 is the wavelength of the plane wave in the medium surrounding the sphere, and λ_1 is its wavelength in the sphere. The sphere radius is a.

Eq. (13b) becomes

$$\psi_s\simeq A_s r^{-1}\{[J_{\frac{1}{2}}(\omega a)/(\omega a)^{\frac{1}{2}}]-\kappa e^{2ik_1a}J_{\frac{1}{2}}(va)/(va)^{\frac{1}{2}}\}, \quad (20a)$$

where

$$A_s=a^3(2\pi)^{\frac{1}{2}}k_1(k_1-k_0)e^{i(k_1-k_0)a+ik_0r}/(1-\kappa^2e^{4ik_1a}). \quad (20b)$$

The differential scattering cross section $\sigma_s(\theta)d\Omega$, or fraction of initial energy scattered through a solid angle $d\Omega$ at θ is given by

$$\sigma_s(\theta)d\Omega=r^2|\psi_s|^2d\Omega$$

$$=|A_s|^2\left\{\frac{J_{\frac{1}{2}}^2(\omega a)}{(\omega a)^3}-2\kappa\cos 2k_1a\frac{J_{\frac{1}{2}}(\omega a)J_{\frac{1}{2}}(va)}{(\omega va^2)^{\frac{1}{2}}}+\kappa^2\frac{J_{\frac{1}{2}}^2(av)}{(va)^3}\right\}d\Omega, \quad (21a)$$

$$|A_s|^2=2\pi a^6k_1^2(k_1-k_0)^2/(1+\kappa^4-2\kappa^2\cos 4k_1a). \quad (21b)$$

[18] Watson, *Bessel Functions* (Cambridge University Press, London, 1948), second edition, p. 363.

In the limit as $k_1\to k_0$, the differential scattering cross section becomes

$$\sigma_s(\theta)\sim 2\pi a^6k_0^2(k_1-k_0)^2J_{\frac{1}{2}}^2(4k_0a\sin\theta)/(4k_0a\sin\tfrac{1}{2}\theta)^3,$$

which agrees with the Rayleigh-Gans-Born approximation.

The total scattering cross sections can be obtained from Eq. (21a) by integrating over all θ. The integration over the cross term $J_{\frac{1}{2}}(\omega a)J_{\frac{1}{2}}(va)$ does not seem to be possible in closed form. However, we can obtain simple results by integrating the first term. We neglect the terms of order κ and κ^2 in both numerator and denominator of Eq. (21) to obtain

$$\sigma_s=2\pi\int_0^{\pi}\sigma_s(\theta)\sin\theta d\theta$$

$$\simeq\frac{2\pi|A_s|^2}{k_1k_0a^2}\int_{\theta=0}^{\theta=\pi}(\omega a)^{-2}J_{\frac{1}{2}}^2(\omega a)d(\omega a)$$

$$=\frac{2\pi|A_s|^2}{k_0k_1a^2}\int_{a(k_1-k_0)}^{a(k_1+k_0)}x^{-2}J_{\frac{1}{2}}^2(x)dx. \quad (22)$$

It is easy to verify by direct application of the Bessel function recursion formulas that

$$(d/dx)\{J_{\frac{1}{2}}^2(x)/x+J_{\frac{1}{2}}^2(x)/x\}=-4J_{\frac{1}{2}}^2(x)/x^2. \quad (23)$$

Hence,

$$\sigma_s\simeq\pi^2a^4k_1(k_1-k_0)^2k_0^{-1}\left\{\frac{J_{\frac{1}{2}}^2[a(k_1-k_0)]+J_{\frac{1}{2}}^2[a(k_1-k_0)]}{a(k_1-k_0)}-\frac{J_{\frac{1}{2}}^2[a(k_1+k_0)]+J_{\frac{1}{2}}^2[a(k_1+k_0)]}{a(k_1+k_0)}\right\}. \quad (24)$$

The function $\sigma_s/\pi a^2$ is plotted in Fig. 1 as a function of $x=ak_0=2\pi a/\lambda_0$ when $m=k_1/k_0=1.1$. Note that as $x\to\infty$, $\sigma_s/\pi a^2$ remains finite. This is a considerable improvement over the RGB cross sections, which go monatonically to infinity. There is still much to be desired, however, because as $x\to\infty$,

$$\sigma_s\to 2\pi a^2\{4/(1+k_0/k_1)^2\}, \quad (25)$$

instead of $2\pi a^2$, the correct value. As the softness of the scatterer increases (i.e., as $k_1/k_0\to 1$), σ_s correctly approaches twice the geometrical cross section.

The internal wave function of the sphere can be determined immediately by solving Eq. (14) for the coefficients $A_{t,n}$ and simplifying them by application of Eqs. (18a) and (18c):

$$A_{t,n}=\frac{2k_1}{(k_1+k_0)}e^{i(k_1-k_0)a}\frac{[1-(-1)^n\kappa e^{2ik_1a}]}{[1-\kappa^2e^{4iak_1}]}.$$

If this expression is substituted into Eq. (13c), the resulting sums are Eqs. (13a) and (13b) with $\cos\theta$

replaced by $-\cos\theta$. Hence,

$$\psi_t=\frac{2k_1e^{i(k_1-k_0)a}[e^{ik_1z}-\kappa e^{2ik_1a}e^{-ik_1z}]}{(k_1+k_0)[1-\kappa^2e^{4iak_1}]}.$$

It is interesting that this internal wave function is very similar to that obtained in an infinite plate of thickness $2a$ (whose center is at the origin and whose surfaces are parallel to the xy plane) and propagation constant k_1 when a plane wave of propagation constant k_0 impinges on the plate. In that case

$$\psi_t=\frac{2k_0e^{ia(k_1-k_0)}[e^{izk_1}+\kappa e^{2iak_1}e^{-izk_1}]}{(k_1+k_0)(1-\kappa^2e^{4ik_1a})}.$$

As is to be expected, the wave function ψ_t reduces to e^{ik_1z} as $k_1\to k_0$.

IV. SCATTERING OF ELECTROMAGNETIC WAVES

We shall apply the methods of the last section to the scattering of plane vector waves. The starting point of this theory is, of course, Maxwell's equations. It was shown by Mie[10] that the solution of the vector wave equation can be expressed as certain vector operators multiplied on solutions of the scalar wave equation.[10, 19, 20] Since the complete derivation of the required equations is readily available in Stratton's book, we shall merely quote the results.

As before, we let a be the radius of our homogeneous isotropic obstacle, λ_0 the wavelength of the incident plane wave in the surrounding medium, and λ_1 its wavelength in the sphere. The ratio of the propagation constants $k_1/k_0(k_j=2\pi/\lambda_j)$ is the relative index of refraction of the sphere, m. The incident beam is chosen to be polarized in the x direction and propagated in the z direction. The origin of the coordinate system is placed at the center of the sphere. As in the Rayleigh-Gans theory (Eq. (7b)), we use the abbreviation $x=2\pi a/\lambda$. The choice of letters x, ρ, and m follows the notation of van de Hulst. Our x, m, and k_0 are equivalent, respectively, to Stratton's ρ, N, and k_2.

The electric field at a point R, θ, φ outside the sphere is given by

$$\mathbf{E}=E_0e^{-i\omega t}\sum_{n=1}^{\infty}i^n\frac{(2n+1)}{n(n+1)}(a_n\mathbf{m}_n-ib_n\mathbf{n}_n), \quad (26a)$$

where

$$\mathbf{m}_n=(1/\sin\theta)h_n(Rk_0)P_n^{(1)}(\cos\theta)\cos\varphi\,\mathbf{i}_2 - h_n(Rk_0)[\partial P_n^{(1)}/\partial\vartheta]\sin\varphi\,\mathbf{i}_3, \quad (26b)$$

$$\mathbf{n}_n=[n(n+1)/Rk_0]h_n(Rk_0)P_n^{(1)}(\cos\theta)\cos\varphi\,\mathbf{i}_1 + (1/Rk_0)[k_0Rh_n(k_0R)]'[\partial P_n^{(1)}/\partial\theta]\cos\varphi\,\mathbf{i}_2 - (1/Rk_0\sin\theta)[Rk_0h_n(Rk_0)]' \times P_n^{(1)}(\cos\theta)\sin\varphi\,\mathbf{i}_3, \quad (26c)$$

$$a_n=-\frac{j_n(mx)[xj_n(x)]'-j_n(x)[mxj_n(mx)]'}{j_n(mx)[xh_n(x)]'-h_n(x)[mxj_n(mx)]'}, \quad (26d)$$

$$b_n=-\frac{j_n(x)[mxj_n(mx)]'-m^2j_n(mx)[xj_n(x)]'}{h_n(x)[mxj_n(mx)]'-m^2j_n(xm)[xh_n(x)]'}. \quad (26e)$$

The unit vectors $\mathbf{i}_1$, $\mathbf{i}_2$, and $\mathbf{i}_3$ are the unit vectors in the R, θ, and φ directions in polar coordinates. $P_n^{(1)}(\cos\theta)$ is the nth associated Legendre polynomial of order one.

We shall be interested in the electric field only at points many sphere diameters away from the scatterer (R/a large). Hence, we can insert the asymptotic expression for $h_n(Rk_0)$ (Eq. (17)) into Eqs. (26b) and (26c) to obtain

$$\mathbf{m}_n\sim\frac{(-i)^{n+1}e^{iRk_0}}{Rk_0}\times\left[\frac{P_n^{(1)}(\cos\theta)}{\sin\theta}\cos\varphi\,\mathbf{i}_2-\frac{\partial P_n^{(1)}}{\partial\theta}\sin\varphi\,\mathbf{i}_3\right]. \quad (27a)$$

The component of $\mathbf{i}_1$ is negligible in $\mathbf{n}_n$. Hence,

$$\mathbf{n}_n\sim\frac{i(-i)^{n+1}e^{iRk_0}}{Rk_0}\times\left[\frac{\partial P_n^{(1)}}{\partial\theta}\cos\varphi\,\mathbf{i}_2-P_n^{(1)}(\cos\theta)\sin\varphi\,\mathbf{i}_3\right]. \quad (27b)$$

In order to sum the series (26a), we shall make approximations in the denominators of a_n, Eq. (26d), and b_n, Eq. (26e), similar to those made in Eqs. (15) and (18a). Again, we successively make the assumptions that $mx>x\gg n$ and $|m-1|<1$. The expressions resulting from the first reduce to those obtained on the basis of the second. Hence, under both of these conditions the denominators of $a_n(D_n{}^a)$ and $b_n(D_n{}^b)$ (which result from substituting Eq. (17) into Eqs. (26d) and (26e) are

$$D_n{}^a\simeq[i(m+1)/2mx]e^{-i(m-1)x}[1+(-1)^n\kappa e^{2imx}], \quad (28a)$$

$$D_n{}^b\simeq[-i(m+1)/2x]e^{-i(m-1)x}[1-(-1)^n\kappa e^{2imx}], \quad (28b)$$

$$\kappa=(k_1-k_0)/(k_1+k_0)=(m-1)/(m+1).$$

The numerators of Eqs. (27a) and (27b) can be simplified considerably by using the well-known recursion formulas for Bessel functions. When one combines the resulting expressions with Eq. (28), he obtains

$$a_n\simeq\frac{\pi(m-1)im^{\frac{1}{2}}e^{ix(m-1)}[1-\kappa(-1)^ne^{2ixm}]}{[1-\kappa^2\exp(4imx)]}\times\int_0^x J_{n+\frac{1}{2}}(\rho m)J_{n+\frac{1}{2}}(\rho)\rho d\rho, \quad (29a)$$

[19] P. Debye, Ann. Physik **30**, 57 (1909).
[20] See reference 16, pp. 392, 554, and 563.

$$b_n \simeq \frac{\pi(m-1)ie^{ix(m-1)}[1+\kappa(-1)^n e^{2ixm}]}{m^{\frac{1}{2}}[1-\kappa^2\exp 4imx]} \times\left\{\int_0^x \rho J_{n+\frac{1}{2}}(m\rho)J_{n+\frac{1}{2}}(\rho)d\rho + x^{\frac{1}{2}}J_{n+\frac{1}{2}}(mx)\frac{d}{dx}(x^{\frac{1}{2}}J_{n+\frac{1}{2}}(x))\right\}. \quad (29b)$$

Since the radial component of E is proportional to R^{-2}, while the others are proportional to R^{-1}, we shall neglect it. From Eqs. (26) and (27) we have

$$E_\theta/|E_0| = \mathbf{E}\cdot\mathbf{i}_2/|E_0| = \frac{i\cos\varphi}{Rk_2}\sum_{n=1}^{\infty}\frac{(2n+1)}{n(n+1)} \times\left[\frac{a_n P_n^{(1)}(\cos\theta)}{\sin\theta}+b_n\frac{\partial P_n^{(1)}}{\partial\theta}\right]. \quad (30a)$$

$$\frac{E_\varphi}{|E_0|} = \frac{i\sin\varphi\, e^{iRk_0}}{Rk_0}\sum_{n=1}^{\infty}\frac{2n+1}{n(n+1)} \times\left[a_n\frac{\partial P_n^{(1)}}{\partial\theta}+b_n\frac{P_n^{(1)}(\cos\theta)}{\sin\theta}\right], \quad (30b)$$

with the approximations (29)

$$\frac{E_\theta}{|E_0|} = \frac{\pi(m-1)\cos\varphi e^{i(m-1)x}e^{iRk_0}}{m^{\frac{1}{2}}Rk_0\sin\theta[1-\kappa^2\exp 4imx]} \times\{F_1(\theta)+\kappa e^{2imx}F_1(\pi-\theta)\}, \quad (31a)$$

$$\frac{E_\varphi}{|E_0|} = -\frac{\pi(m-1)\sin\varphi e^{i(m-1)x}e^{iRk_0}}{m^{\frac{1}{2}}Rk_0\sin\theta[1-\kappa^2\exp 4imx]} \times\{F_2(\theta)-\kappa e^{2imx}F_2(\pi-\theta)\}. \quad (31b)$$

The functions $F_1(\theta)$ and $F_2(\theta)$ are defined by

$$F_1(\theta) = mS_1(\theta)+S_2(\theta)+S_4(\theta), \quad (32a)$$

$$F_2(\theta) = mS_2(\theta)+S_1(\theta)+S_3(\theta), \quad (32b)$$

where the $S_j(\theta)$'s represent the sums

$$\begin{Bmatrix}S_1(\theta)\\ S_2(\theta)\end{Bmatrix} = \sum_{n=1}^{\infty}\begin{Bmatrix}P_n^{(1)}(\cos\theta)\\ \sin\theta[\partial P_n^{(1)}/\partial\theta]\end{Bmatrix}\frac{(2n+1)}{n(n+1)} \times\int_0^x zJ_{n+\frac{1}{2}}(mz)J_{n+\frac{1}{2}}(z)dz, \quad (33a)$$

$$\begin{Bmatrix}S_3(\theta)\\ S_4(\theta)\end{Bmatrix} = \sum_{n=1}^{\infty}\begin{Bmatrix}P_n^{(1)}(\cos\theta)\\ \sin\theta[\partial P_n^{(1)}/\partial\theta]\end{Bmatrix}\frac{(2n+1)}{n(n+1)} \times x^{\frac{1}{2}}J_{n+\frac{1}{2}}(mx)\frac{d}{dx}[x^{\frac{1}{2}}J_{n+\frac{1}{2}}(x)]. \quad (33b)$$

In deriving Eqs. (31a) and (31b) the identity $P_n^{(1)}(\cos\theta) = (-1)^{n+1}P_n^{(1)}[\cos(\pi-\theta)]$ was used. The sums $S_j(\theta)$ and $S_j(\pi-\theta)$, which are closely related to Eq. (19), are evaluated in the appendix. It is there shown that

$$\begin{Bmatrix}F_1(\theta)\\ F_2(\theta)\end{Bmatrix} = \left(\frac{2m}{\pi}\right)^{\frac{1}{2}}\frac{mJ_{\frac{1}{2}}(\omega x)\sin\theta}{(\omega/x)^{\frac{1}{2}}}\begin{Bmatrix}\cos\theta\\ 1\end{Bmatrix}, \quad (34)$$

where $\omega^2 = 1+m^2-2m\cos\theta$. In the corresponding expression for $F_j(\pi-\theta)$, ω is replaced by

$$v = (m^2+1+2m\cos\theta)^{\frac{1}{2}}.$$

The relative intensities $I_\theta/I_0 = |E_\theta/E_0|^2$ and $I_\varphi/I_0 = |E_\varphi/E_0|^2$ are given by

$$\begin{Bmatrix}I_\theta/I_0\\ I_\varphi/I_0\end{Bmatrix} = \frac{2\pi(m-1)^2a^2x^4m^2}{R^2(1+\kappa^4-2\kappa^2\cos 4mx)}\begin{Bmatrix}\cos^2\theta\cos^2\varphi\\ \sin^2\varphi\end{Bmatrix} \times\left|\frac{J_{\frac{1}{2}}(\omega x)}{(\omega x)^{\frac{1}{2}}}-\kappa\frac{e^{2imx}J_{\frac{1}{2}}(vx)}{(vx)^{\frac{1}{2}}}\right|^2. \quad (35a)$$

This pair of equations is equivalent to the statement that the intensity measured with an analyzer oriented in a direction $\mathbf{q}$ is

$$I_q/I_0 = \frac{2(\mathbf{p}\cdot\mathbf{q})^2\pi(m-1)^2a^2m^2x^4}{R^2(1+\kappa^4-2\kappa^2\cos 4mx)} \times\left|\frac{J_{\frac{1}{2}}(\omega x)}{(\omega x)^{\frac{1}{2}}}-\kappa\frac{e^{2ixm}J_{\frac{1}{2}}(vx)}{(vx)^{\frac{1}{2}}}\right|^2, \quad (35b)$$

where $\mathbf{p}$ is a unit vector in the direction of polarization of the incident radiation. The greatest weakness of this result is that E_θ and E_φ show no phase difference. Hence to our approximation, scattered light is linearly polarized. As $m\to 1(\kappa\to 0)$, Eq. (35) reduces to the Rayleigh-Gans expression.

The total scattering cross section, σ_s, can be computed from Eq. (9). We shall show that $\sigma_s/\pi a^2$ remains finite as $x=2\pi a/\lambda\to\infty$, and indeed, as $m\to 1$, that it correctly approaches two. As in the scalar case, we shall limit ourselves to the range $m\leqslant 1.5$. Then to a first approximation (which becomes increasingly better as $m\to 1$) the terms of order κ and κ^2 can be dropped in Eq. (35). By substituting Eq. (35) into Eq. (9) we obtain

$$\sigma_s \simeq 2m^2\pi^2(m-1)^2a^2x^4 \times\int_0^\pi (\omega x)^{-3}J_{\frac{1}{2}}^2(\omega x)(1+\cos^2\theta)\sin\theta d\theta. \quad (36)$$

If, as in the scalar case, we introduce the new variable $z=\omega x$, we find

$$\sigma_s \simeq [\pi^2a^2(m-1)^2x^2/2m]\int_{x(m-1)}^{x(m+1)} J_{\frac{1}{2}}^2(z) \times\{(m^4+6m^2+1)z^{-2}-2(m^2+1)x^{-2}+z^2x^{-4}\}dz.$$

The required integrals have been recorded in Eqs. (10a)–(10d). By using the notation there adopted, we see that

$$\sigma_s \simeq [\pi^2 a^2 x^2 (m-1)^2/2m]\{(m^4+6m^2+1)\Delta_1 - 2(m^2+1)x^{-2}\Delta_2 + x^{-4}\Delta_3\}, \quad (37)$$
$$\Delta_j = I_j(x[m+1]) - I_j(x[m-1]).$$

As $x\to\infty$, the limiting value for $\sigma_s/\pi a^2$ can be found by examining the asymptotic behavior of the integrals I_1, I_2, and I_3. The function $\varphi(z)$ (Eq. (10d)) has the property that $\varphi(z_2)-\varphi(z_1)\sim\log(z_2/z_1)$ as z_1 and z_2 become infinite. Also for large z, $[J_{\frac{1}{2}}^2(z)+J_{\frac{1}{2}}^2(z)]z^{-1}\sim 2/\pi z^2$.

Hence, as $x\to\infty$,

$$\Delta_1\sim 2m/\pi x^2(m^2-1)^2,$$
$$\Delta_2\sim \pi^{-1}\log[(m+1)/(m-1)],$$
$$\Delta_3\sim 2mx^2/\pi,$$

so that

$$\sigma_s/\pi a^2 \to 2\frac{(m^2+1)^2}{(m+1)^2}\left\{1-\frac{(m^2-1)^2}{2m(m^2+1)}\log\left(\frac{m+1}{m-1}\right)\right\}. \quad (38)$$

As $m\to 1$, $\sigma_s/\pi a^2\to 2$ as is desired. When $m=1.1$, 1.25, 1.33, and 1.50, this ratio becomes 2.15, 2.31, 2.38, 2.51, so that our cross sections for large spheres will be somewhat high. These might be improved by taking into account the terms of order κ and κ^2 in Eq. (35b).

In the limit as $m\to 1$, a natural scattering parameter is[6] $\rho=2x(m-1)$. Then, for all fixed ρ, $x(m+1)=\frac{1}{2}\rho(m+1)/(m-1)\to\infty$, and

$$\sigma_s/\pi a^2 \to \tfrac{1}{2}\pi\rho\{J_{\frac{1}{2}}^2(\tfrac{1}{2}\rho)+J_{\frac{1}{2}}^2(\tfrac{1}{2}\rho)\} = 2-4\rho^{-1}\sin\rho+4\rho^{-2}(1-\cos\rho). \quad (39)$$

This exact limiting formula for the extinction has also been derived in a completely different manner by Van de Hulst[6] (Eq. (8.6) of his paper I). As long as $(m-1)\ll 1$ (or more precisely when $m<1.1$), our approximations to the coefficients of the Mie series are excellent, and Eq. (37) yields values for the extinction $E=\sigma_s/\pi a^2$ which deviate from the exact ones by at most about 5 percent for all positive x.

It is to be noticed (Fig. 2) that in the limit, as $m\to 1$, the first maximum in the extinction appears at $x\simeq 2.08/(m-1)$. Indeed, Fig. 3 and auxiliary calculation show that this is still true for all $m\leq 1.5$.

The extinction is plotted in Fig. 2 as a function of $\frac{1}{2}\rho$ when $m=1$ and $m=1.25$. It is plotted in Fig. 3 as a function of x when $m=1.33$ and 1.50. These curves compare very well with the exact ones obtained by numerical summation of the Mie series by the Bureau of Standards Computation Laboratory.[7, 11] To obtain a better idea of the range of validity of Eq. (37) we have tabulated the exact results of the Mie theory for $m=1.33$ and 1.5 and compared them with our approximations (Tables I and II).

Tables I and II indicate deviations of less than 5 percent when the extinction is greater than 1. Of course, the deviations eventually exceed this level when x gets sufficiently large because in the limit, $E=2.38$ when $m=4/3$, and $E=2.51$ when $m=1.5$, rather than the correct value, $E=2$.

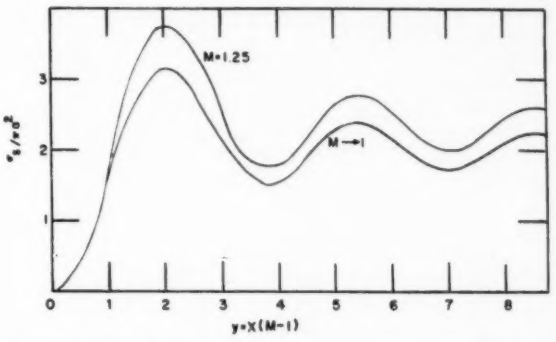


FIG. 2. Ratio of scattering cross section to sphere cross section for an electromagnetic wave as a function of $y=2\pi a(m-1)/\lambda_0$, where m is the relative index of scatterer to surrounding medium.

The fact that our formula for the extinction is much better in the range of intermediate values of x than it is for very small or very large values of x can be explained qualitatively on the basis of our approximations (28) of the denominators of the Mie coefficients. Physically, each term in the Mie series corresponds to the contribution of a particular normal mode of the excited sphere to the total scattered field at a given point. As $x\to 0$, the lowest mode (first term in the Mie series) predominates over the others. As x increases, more and more modes become significant. Indeed, it was shown by Debye[19] that most of the contribution to the extinction comes from the first n_0 terms in the Mie series where $n_0\simeq x$. When $(m-1)$ is not $\ll 1$, our approximation of the denominators of the Mie coefficients was based on the hypothesis that $x\gg n$. The asymptotic formulas[17] that we used for the Bessel functions bear a slight resemblance to these functions, even when $x\simeq n$ when x is large; however, when $n>x$ even this disappears. When x is very small, say $x\ll 1$, there is no coefficient whose subscript n is less than x. Hence, our approximations to the coefficients are poor for every n and our extinctions can be expected to be inaccurate. This is borne out in Tables I and II.

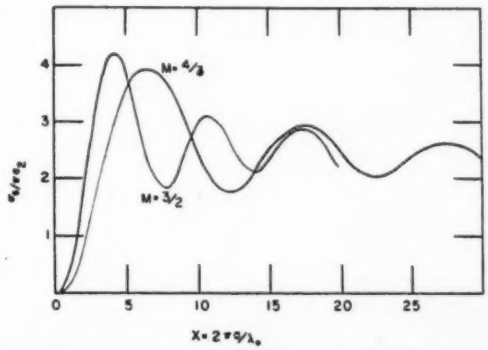


FIG. 3. Ratio of scattering cross section to sphere cross section as a function of $x=2\pi a/\lambda_0$ for relative index of refraction 4/3 and 3/2.

TABLE I. Ratio of total scattering cross section to geometrical cross section. Comparison of exact values (reference 11) with results of Eq. (37). $m=k_1/k_0=4/3$, a=radius of scatterer, $x=2\pi a/\lambda_0=ak_0$, where λ_0 is wavelength of incident radiation.

x	$\sigma_s/\pi a^2$ exact	$\sigma_s/\pi a^2$, Eq. (37)
0.6	0.014	0.025
1.0	0.094	0.147
1.2	0.171	0.239
1.5	0.322	0.410
1.8	0.523	0.616
2.0	0.710	0.773
2.4	1.13	1.15
2.5	1.21	1.26
3.0	1.75	1.79
3.6	2.38	2.42
4.0	2.82	2.82
4.8	3.49	3.46
5.0	3.59	3.58
6.0	3.89	3.91

When x is near $2.08/(m-1)$, the location of the first peak in the extinction curve n_0 varies from 20 to 4 as m varies from 1.1 to 1.5. In this range n_0 is sufficiently large so that $x \gg n$ for some of the important terms and certainly $x>n$ for all of them. Since our Mie coefficients are quite accurate in this range, we expect accurate values of the extinction, as was observed in Tables I and II. Since $n_0=4$ for $m=1.5$ is somewhat small for our described principle to apply, it is not surprising that we did not obtain the existing $m=1.5$ fine structure in the first peak.

TABLE II. Ratio of total scattering cross section to geometrical cross section. $m=\frac{3}{2}$.

$x=2\pi a/\lambda_0$	$\sigma_s/\pi a^2$ exact	$\sigma_s/\pi a^2$ (Eq. 37)
0.5	0.015	0.037
0.6	0.030	0.069
1.0	0.215	0.349
1.2	0.395	0.592
1.5	0.753	0.993
1.8	1.35	1.43
2.0	1.80	1.76
2.4	2.34	2.48
2.5	2.54	2.67
3.0	3.42	3.43
3.2	3.53	3.66
3.4	3.88	3.86
3.6	4.18	4.02
3.8	4.13	4.13
4.0	4.05	4.21
4.3	4.36	4.19
4.5	4.20	4.14
4.8	3.82	3.96
5.0	3.93	3.82
5.5	3.18	3.35
6.0	2.90	2.80
6.5	2.37	2.33
7.0	1.85	1.98
7.2	1.77	1.89
8.0	1.78	1.87
8.4	2.01	2.01
9.0	2.24	2.36
9.6	2.80	2.73
10.0	2.88	2.89
10.8	2.98	3.07
12.0	2.48	2.77

As x gets very large, the scattered field at a given point is the resultant of the contribution of a very large number of normal modes. Since there is considerable destructive interference between these components, the total field may be small even though many of the components may be large. Hence, many small errors in each component can accumulate to give a large relative error in the small total field. On this basis our results for very large x can be expected to be in error.

The approximation of the extinction of small spheres can be improved by using all the terms in Eq. (35b) rather than only the first. In the limit of infinitesimal spheres ($x\to 0$), the value of $\sigma_s/\pi a^2$ obtained from Eq. (35b) with terms of order κ and κ^2 neglected is $\sigma_s/\pi a^2 \sim 32m^2(m-1)^2x^4/27$, which is essentially the Rayleigh-Gans relation. When all terms of Eq. (35b) are included, the corresponding expression becomes $\sigma_s/\pi a^2 \sim 8x^4(m^2-1)^2/27$. This differs from the first term of the exact formula,

$$\frac{\sigma_s}{\pi a^2}=\frac{8}{3}x^4\left(\frac{m^2-1}{m^2+2}\right)^2\left[1+\frac{6}{5}\left(\frac{m^2-2}{m^2+2}\right)x^4+\cdots\right] \quad (40)$$

by having the quantity (m^2+2) replaced by 3.

By expressing the extinction by Eq. (40) when it is less than 1 and then switching to Eq. (37) when it exceeds 1, one has relatively simple analytical representations whose errors in their proper range are about 5 percent or less through the first few peaks, provided that $m<1.5$. When $m=1.5$ or larger, the terms of order κ and κ^2 (which we have neglected) in the extinction formula become important. It is to be noted that, although these terms improve the asymptotic results as $x\to 0$, they do not contribute significantly in the neighborhood of the first peak when $m<1.5$. Their contribution is small because $\kappa^2<1/25$ and the only term with a coefficient $\kappa(<1/5)$ is multiplied by the integral

$$\int_0^\pi J_{\frac{1}{2}}(\omega x)J_{\frac{1}{2}}(vx)(1+\cos^2\theta)\sin\theta(v\omega x^2)^{-\frac{1}{2}}d\theta.$$

When x is large, the oscillations of the Bessel functions are out of phase, so there is considerable cancellation of the contributions of the integrand in various parts of the range of integration $0\leq\theta\leq\pi$. Since the term in the extinction that is not multiplied by κ or κ^2 involves the square of a single Bessel function, it is everywhere positive, and is not reduced by the cancelling effect. Since in small spheres less than one cycle of the Bessel functions contributes to the integral, this cancelling does not reduce the contribution of the κ-term.

The range of validity of Eq. (35b) for scattering as a function of angle is, paradoxically, different from that of the extinction. In the case of small scatterers the ratio of $I(\theta)/I(0)$ is in very good agreement with the exact results (Fig. 4) even though $I(\theta)$ is not. This is a

property of the Rayleigh-Gans theory as well as ours. It arises from the fact that in both theories the constant multiplier, a, if we write $I(\theta)=af(\theta)$, is less accurate than $f(\theta)$. Since most scattering measurements are expressed in terms of ratios, it is fortunate that the large error takes place in an unimportant term. Unfortunately, in large spheres for which our extinction values are excellent, our scattering as a function of angle is not so good. This is a consequence of the observation that the extinction is very simply related to the scattering in the forward direction only (see van de Hulst,[6] p. 56). Hence, our theory need only give good results for $I(0)$ (as it does for medium large spheres) to give good value of the extinction. When the intensity is small at an angle θ when compared with forward scattering ($\theta=0$) (say, less than 10 percent), our intensity values are in error but in general by only half that of the corresponding Rayleigh-Gans results (see Fig. 5).

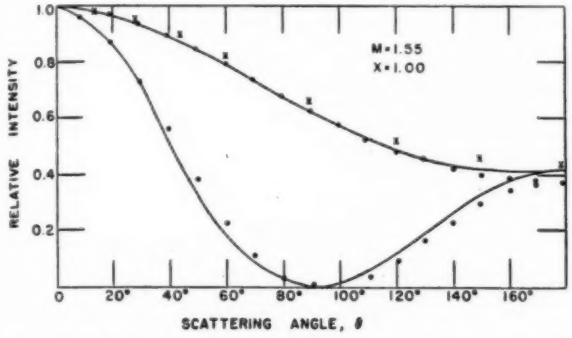


FIG. 4. Relative intensity of scattered light (i.e., ratio of intensity of scattered light at an angle θ to the direction of the initial beam to that of light scattered in the direction of the initial beam) as a function of θ when the relative index $m=1.55$ and the size parameter $x=2\pi a/\lambda_0=1$. The solid line corresponds to our approximate theory, the dots to the Mie theory, and the crosses to the Rayleigh-Gans approximation. The upper curve corresponds to the component of the electric vector perpendicular to the plane formed by the direction of scattered beam and the initial beam. The lower curve corresponds to the component normal to the above and to the direction of propagation of the scattered wave.

The Rayleigh-Gans theory has been used to determine sphere sizes by making a plot of the ratio of intensity of light scattered at two selected angles as a function of the size parameter, x. When the ratio is determined experimentally, the size of the sphere is read off the "dissymmetry curve."[4, 14] Application of the Rayleigh-Gans theory outside its limited range of validity $x\simeq[4\pi(m-1)]^{-1}$ has led to confusion, however, because measurements of ratios at several angles (or wavelengths) yield different theoretical sizes. Figure 6 shows dissymmetry ratios for several angles, $m=4/3$. Again it appears that the approximate theory gives the ratios correct to within about 5 percent so long as the ratio is not less than 0.1. One may always select the angles at which the ratios are taken in order to satisfy this condition. A further condition for validity of the theory must be emphasized.

Experimental measurement of scattering by many

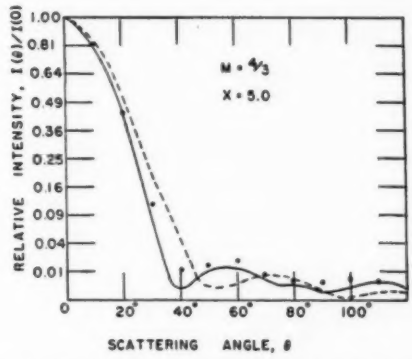


FIG. 5. Relative intensity as a function of angle for component of scattered light whose electric vector is perpendicular to the plane formed by the direction of scattered beam and initial beam. Solid line corresponds to our approximation, broken line to Rayleigh-Gans, and dots to exact Mie theory.

(noninteracting) spheres will not yield a single size independent of wavelength and choice of dissymmetry angles unless the particles are mono-size. A relatively few large particles (for example, dust particles passed through a fine filter) with sizes of the order $x=10$ cause a relatively large increase in small angle scattering. This effect may be minimized by ignoring measurements at small scattering angles.

The theory developed in this paper for the scattering by spheres can be extended to scattering by cylinders. The detailed results for that case will be discussed in a future report.

The authors wish to thank the Research Corporation for the grant which supported this work.

APPENDIX. EVALUATION OF SUMS IN EQ. (33)

The starting point in the evaluation of the required sums is the equivalent of the identity (19a):

$$2(zZ)^{\frac{1}{2}}\pi^{-1}[u^{-1}\sin u-\sin z\sin Z]$$
$$=\sum_{n=1}^{\infty}(2n+1)J_{n+\frac{1}{2}}(z)J_{n+\frac{1}{2}}(Z)P_n^{(0)}(\cos\theta), \quad \text{(A-1)}$$
$$u^2=Z^2+z^2-2zZ\cos\theta.$$

The sum

$$S_0(\theta)=\sum_{n=1}^{\infty}(2n+1)P_n^{(0)}(\cos\theta)\int_0^x xJ_{n+\frac{1}{2}}(xm)J_{n+\frac{1}{2}}(x)dx \quad \text{(A-2)}$$

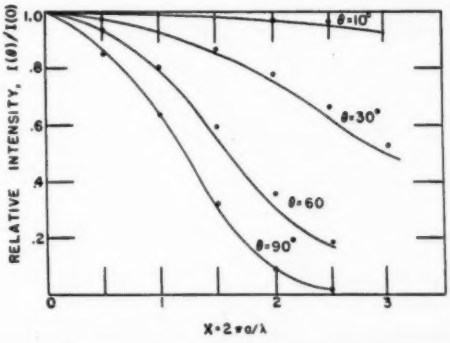


FIG. 6. Relative intensity as a function of x for several scattering angles. These curves correspond to the same component of scattered field as those in Fig. 5. The dots correspond to the exact Mie theory.

follows from Eq. (A-1) by replacing z by x and Z by mx, and integrating from $x=0$ to $x=z$. Then

$$S_0(\theta)=2m^{-\frac{1}{2}}\pi^{-1}\{m[(\sin\omega z-\omega z\cos\omega z)/\omega^3]-f(z)\}, \quad \text{(A-3a)}$$

where

$$f(z)=\int_0^z \sin mx\sin x dx=\frac{1}{2}\left[\frac{\sin(m-1)x}{m-1}-\frac{\sin(m+1)x}{m+1}\right] \quad \text{(A-3b)}$$

and $\omega^2=1+m^2-2m\cos\theta$.

The recursion formulas for legendre polynomials

$$\sin\theta P_n^{(1)}(\cos\theta)/n(n+1)=\int_0^\theta P_n^{(0)}(\cos\theta)\sin\theta d\theta \quad \text{(A-4)}$$

can be applied to Eq. (A-2) to obtain $S_1(\theta)$ (Eq. 33a):

$$S_1(\theta)=\frac{2}{\pi m^{\frac{1}{2}}\sin\theta}\left\{\frac{\sin z(m-1)}{m-1}-\frac{\sin\omega z}{\omega}-f(z)(1-\cos\theta)\right\}. \quad \text{(A-5)}$$

It is obvious from Eq. (33a) that $S_2(\theta)=\sin\theta\partial S_1/\partial\theta$. Hence

$$S_2(\theta)=-S_1(\theta)\cos\theta+\frac{2\sin\theta}{\pi m^{\frac{1}{2}}}\left[\left(\frac{\pi}{2z\omega}\right)^{\frac{1}{2}}\left(\frac{mz^2}{\omega}\right)J_{\frac{3}{2}}(\omega z)-f(z)\right]. \quad \text{(A-6)}$$

To derive an expression for $S_3(\theta)$, let Eq. (A-1) be multiplied by $z^{\frac{1}{2}}$ and the result differentiated with respect to z. The application of Eq. (A-4) and substitutions $z=x$ and $Z=mx$ finally yield

$$S_3(\theta)=\frac{2}{\pi m^{\frac{1}{2}}\sin\theta}\left\{\frac{\sin\omega x}{\omega}(1-m\cos\theta)+\sin(m-1)x-(1-\cos\theta)\cos x\sin mx\right\}. \quad \text{(A-7)}$$

Equation (33b) implies that $S_4(\theta)=\sin\theta\partial S_3(\theta)/\partial\theta$. Hence,

$$S_4(\theta)=-\cos\theta S_3(\theta)+\frac{2}{\pi m^{\frac{1}{2}}}\left\{\frac{m}{\omega}\sin\theta\sin\omega x-\frac{mx^2\sin\theta}{\omega}\left(\frac{\pi}{2\omega x}\right)^{\frac{1}{2}}(1-\cos\theta)J_{\frac{3}{2}}(\omega x)-\sin\theta\sin mx\cos x\right\}. \quad \text{(A-8)}$$

It is merely an elementary exercise in algebraic manipulation to compute the functions $F_1(\theta)$ and $F_2(\theta)$ [Eq. (32)] from Eqs. (A-5)–(A-8) to obtain Eq. (34).

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 4 APRIL, 1951

Separation of Variables in Electromagnetic Theory*

DOMINA EBERLE SPENCER
University of Connecticut, Storrs, Connecticut

(Received November 13, 1950)

Separability conditions are obtained for the partial differential equations of electromagnetic theory. Wave guide and antenna problems are expressed in terms of the vector Helmholtz equation, and solutions are indicated by use of the simple method of separation of variables without recourse to Green's functions. It is shown that complete separation occurs only in rectangular coordinates, but that separation of the equation for one component of the vector field occurs in two coordinate systems. With cylindrical or rotational symmetry, separation is possible in eight coordinate systems.

I. INTRODUCTION

THE method of separation of variables is of such wide usefulness in the solution of problems which can be described in terms of a scalar potential that it is important to investigate just how far this method can be extended in the more complex problems of electromagnetic theory. The main difficulty is that the differential equation now contains the vector laplacian $\nabla^2\mathbf{A}$ instead of the scalar laplacian $\nabla^2\varphi$. And, although the two happen to have the same form in rectangular coordinates, their laws of transformation are different and their expressions in generalized coordinates are by no means identical. Since the expression $\nabla^2\mathbf{A}$ is much more complex than $\nabla^2\varphi$, one would expect to obtain separability of the vector equation in a smaller number of coordinate systems than for the corresponding scalar equation.

It is well known that separability methods can be used in rectangular coordinates, but no thorough investigation seems to have been made of what other coordinate systems can be used. Morse and Feshbach[1] give an elaborate discussion in terms of Green's functions, but come to no definite conclusions on how many coordinate systems are useful in this method. The purpose of this paper is to take the more direct and simple method of separation of variables and to try to extend the procedures[2,3] used with the scalar Laplace and Helmholtz equations to the vector Helmholtz equation. Separability conditions will be derived, and cases in which separation of variables is successful will be enumerated.

The problems of electromagnetic theory can be reduced to the following form, which is especially convenient for wave guide or antenna problems. Write the Maxwell equations as

$$\begin{aligned} &\text{div}\mathbf{D}=\rho, \quad \text{curl}\mathbf{E}=-\partial\mathbf{B}/\partial t,\\ &\text{div}\mathbf{B}=0, \quad \text{curl}\mathbf{H}=\mathbf{J}+\partial\mathbf{D}/\partial t, \end{aligned} \quad (1)$$

and assume a homogeneous isotropic medium and a charge-free space:

$$\begin{aligned} &\mathbf{D}=\epsilon\mathbf{E}, \quad \mathbf{B}=\mu\mathbf{H}, \quad \mathbf{J}=\sigma\mathbf{E}, \quad \rho=0,\\ &\mathbf{E}=\mathbf{E}_0(u^1,u^2,u^3)e^{i\omega t}, \quad \mathbf{H}=\mathbf{H}_0(u^1,u^2,u^3)e^{i\omega t}. \end{aligned} \quad (2)$$

* Presented to the American Mathematical Society, Session on Applied Mathematics, New York City, October 28, 1950.

[1] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (Technology Press, Cambridge, Massachusetts, 1946).

[2] H. P. Robertson, Math. Ann. 98, 749 (1927).

[3] P. Moon and D. E. Spencer, "Separability conditions for the Laplace and Helmholtz equations" (to be published).

Then the Maxwell equations become

$$\begin{aligned} &\mathrm{div}\mathbf{E}_0=0, \quad \mathrm{curl}\mathbf{E}_0=-i\omega\mu\mathbf{H}_0, \\ &\mathrm{div}\mathbf{H}_0=0, \quad \mathrm{curl}\mathbf{H}_0=(\sigma+i\omega\epsilon)\mathbf{E}_0. \end{aligned} \tag{3}$$

The differential equations satisfied by E_0 and H_0 are

$$\nabla^2\mathbf{E}_0=\Gamma^2\mathbf{E}_0, \quad \nabla^2\mathbf{H}_0=\Gamma^2\mathbf{H}_0, \tag{4}$$

where

$$\nabla^2\mathbf{A}=\mathrm{grad}\ \mathrm{div}\mathbf{A}-\mathrm{curl}\ \mathrm{curl}\mathbf{A} \tag{5}$$

and

$$\Gamma^2=i\omega\mu(\sigma+i\omega\epsilon). \tag{6}$$

We consider, therefore, the separability of the vector Helmholtz equation

$$\nabla^2\mathbf{A}+k^2\mathbf{A}=0, \tag{7}$$

where

$$\mathbf{A}=\mathbf{A}(u^1, u^2, u^3).$$

II. CONDITIONS FOR COMPLETE SEPARABILITY

The vector Helmholtz equation in generalized coordinates has the form,

$$\begin{aligned} &(1/g_{ii})^{\frac{1}{2}}(\partial/\partial u^i)[(1/g)^{\frac{1}{2}}(\partial/\partial u^i)(g/g_{ii})^{\frac{1}{2}}(A)_i] \\ &\quad+(g_{ii}/g)^{\frac{1}{2}}(\partial/\partial u^j)[(g_{kk}/g^{\frac{1}{2}})(\partial/\partial u^j)(g_{ii})^{\frac{1}{2}}(A)_i] \\ &\quad+(g_{ii}/g)^{\frac{1}{2}}(\partial/\partial u^k)[(g_{jj}/g^{\frac{1}{2}})(\partial/\partial u^k)(g_{ii})^{\frac{1}{2}}(A)_i] \\ &\quad+(1/g_{ii})^{\frac{1}{2}}(\partial/\partial u^i)[(1/g)^{\frac{1}{2}}(\partial/\partial u^i)(g/g_{jj})^{\frac{1}{2}}(A)_j] \\ &\quad+(1/g_{ii})^{\frac{1}{2}}(\partial/\partial u^i)[(1/g)^{\frac{1}{2}}(\partial/\partial u^k)(g/g_{kk})^{\frac{1}{2}}(A)_k] \\ &\quad-(g_{ii}/g)^{\frac{1}{2}}(\partial/\partial u^j)[(g_{kk}/g^{\frac{1}{2}})(\partial/\partial u^i)(g_{jj})^{\frac{1}{2}}(A)_j] \\ &\quad-(g_{ii}/g)^{\frac{1}{2}}(\partial/\partial u^k)[(g_{jj}/g^{\frac{1}{2}})(\partial/\partial u^i)(g_{kk})^{\frac{1}{2}}(A)_k] \\ &\qquad\qquad+k^2(A)_i=0, \end{aligned} \tag{8}$$

where $i, j, k=1, 2, 3$; $i\neq j\neq k$.

Equation (8) contains two sources of difficulty. In the first place, this equation contains terms in $(A)_1$, $(A)_2$, and $(A)_3$. If the separation method is to be applicable, it is necessary that the terms in $(A)_j$ and $(A)_k$ drop out. Thus, a necessary but not sufficient condition for separability is

$$\begin{aligned} &(\partial/\partial u^i)[(1/g)^{\frac{1}{2}}(\partial/\partial u^j)(g/g_{jj})^{\frac{1}{2}}(A)_j] \\ &\quad=(g_{ii}/g^{\frac{1}{2}})(\partial/\partial u^j)[(g_{kk}/g^{\frac{1}{2}})(\partial/\partial u^i)(g_{jj})^{\frac{1}{2}}(A)_j], \end{aligned} \tag{9}$$

where $i, j, k=1, 2, 3$, $i\neq j\neq k$. If Eq. (9) is satisfied, then Eq. (8) reduces to

$$\begin{aligned} &(1/g_{ii})^{\frac{1}{2}}(\partial/\partial u^i)[(1/g)^{\frac{1}{2}}(\partial/\partial u^i)(g/g_{ii})^{\frac{1}{2}}(A)_i] \\ &\quad+(g_{ii}/g)^{\frac{1}{2}}\{(\partial/\partial u^j)[(g_{kk}/g^{\frac{1}{2}})(\partial/\partial u^j)(g_{ii})^{\frac{1}{2}}(A)_i] \\ &\quad+(\partial/\partial u^k)[(g_{jj}/g^{\frac{1}{2}})(\partial/\partial u^k)(g_{ii})^{\frac{1}{2}}(A)_i]\} \\ &\qquad\qquad+k^2(A)_i=0. \end{aligned} \tag{10}$$

In order that Eq. (10) separate, it is necessary that

$$\begin{aligned} (g_{11})^{\frac{1}{2}}&=f_1(u^1)f_2(u^2)f_3(u^3), \\ (g_{22})^{\frac{1}{2}}&=g_1(u^1)g_2(u^2)g_3(u^3), \\ (g_{33})^{\frac{1}{2}}&=h_1(u^1)h_2(u^2)h_3(u^3), \end{aligned} \tag{11}$$

since it must be possible to separate the functions of u^j and u^k from expressions of the type $\partial(g_{jj})^{\frac{1}{2}}(A)_j/\partial u^i$. It is here that a striking difference between the vector and scalar equations occurs: with the vector Helmholtz equation, terms of the two types

$$\partial[(g_{jj})^{\frac{1}{2}}(A_j)]/\partial u^i \quad \text{and} \quad (\partial/\partial u^i)(g/g_{ii})^{\frac{1}{2}}(A)_i$$

appear, and both must be separable. Thus, a more stringent limitation is imposed on the form of the g_{ii} than in the corresponding scalar wave equation.

Let us see in what coordinate systems the necessary conditions of Eqs. (9) and (11) are satisfied. Since the conditions on the metric coefficients are more severe for the vector Helmholtz equation than for the scalar Helmholtz equation, it appears unlikely that the vector equation will separate for any coordinate system in which the scalar one does not separate. Consideration of the metric coefficients of the eleven coordinate systems[4] in which the scalar Helmholtz equation is known to be separable shows that Eq. (11) is satisfied in only three of these coordinate systems: rectangular, circular-cylinder, and spherical.

Another condition remains to be satisfied: that expressed by the six Eqs. (9). In rectangular coordinates the six equations are satisfied identically. In circular-cylinder coordinates, they become

$$\begin{aligned} &\text{for equation in } A_r, \ \partial A_\psi/\partial\psi=0; \\ &\text{for equation in } A_\psi, \ \partial A_r/\partial\psi=0. \end{aligned}$$

The equation in A_z is always separable. Separability for all three equations is possible only if A_r and A_ψ are independent of ψ. In spherical coordinates, we have from Eqs. (9),

$$\begin{aligned} &\text{for equation in } A_r, \ (\partial A_\theta/\partial\theta)+\cot\theta A_\theta=0, \ \partial A_\psi/\partial\psi=0; \\ &\text{for equation in } A_\theta, \ \partial A_r/\partial\theta=0, \ \partial A_\psi/\partial\psi=0; \\ &\text{for equation in } A_\psi, \ \partial A_r/\partial\psi=0, \ \partial A_\theta/\partial\psi=0. \end{aligned}$$

Theorem I.—The only one of the eleven Eisenhart systems in which the vector Helmholtz equation separates completely is rectangular coordinates.

Theorem II.—The only coordinate system in which one of the three vector Helmholtz equations separates in the general case $\mathbf{A}=\mathbf{A}(u^1, u^2, u^3)$ is circular-cylinder coordinates. In this case, the equation in A_z is always separable.

III. SOLUTION IN CIRCULAR-CYLINDER COORDINATES

As we have seen in Theorem II, only one of the three scalar equations of the vector Helmholtz equation is separable in circular-cylinder coordinates. It turns out, however, that a neat solution of the electromagnetic field problem is possible in this case. The partial differential equations are

$$\nabla^2\mathbf{E}_0=\Gamma^2\mathbf{E}_0 \quad \text{and} \quad \nabla^2\mathbf{H}_0=\Gamma^2\mathbf{H}_0.$$

By Theorem II, the differential equations in E_{0z} and H_{0z} are separable and their solution can be written

$$\begin{aligned} E_{0z}=\sum_{n=0}^{\infty}&[A_nJ_n(qr)+B_nY_n(qr)](\sin n\psi+C_n\cos n\psi) \\ &\times\exp[\pm(q^2-\Gamma^2z)^{\frac{1}{2}}], \end{aligned}$$

$$\begin{aligned} H_{0z}=\sum_{n=0}^{\infty}&[\mathfrak{A}_nJ_n(qr)+\mathfrak{B}_nY_n(qr)](\sin n\psi+\mathfrak{C}_n\cos n\psi) \\ &\times\exp[\pm(q^2-\Gamma^2z)^{\frac{1}{2}}]. \end{aligned}$$

[4] L. P. Eisenhart, Ann. Math. 35, 284 (1934).

TABLE I. Separation of the vector Helmholtz equation.

$\nabla^2\mathbf{A}+k^2\mathbf{A}=0$

1. *Rectangular coordinates.* $\mathbf{A}=\mathbf{A}(x, y, z)$

A_x, A_y, or $A_z=X(x)Y(y)Z(z)$
$X=C\exp[(p^2+q^2)^{\frac{1}{2}}x]+D\exp[-(p^2+q^2)^{\frac{1}{2}}x]$
$Y=C\sin py+D\cos py$
$Z=C\sin[(k^2+q^2)^{\frac{1}{2}}z]+D\cos[(k^2+q^2)^{\frac{1}{2}}z]$

2. *Circular-cylinder coordinates.* General case $\mathbf{A}=\mathbf{A}(r, \psi, z)$

$A_z=R(r)\Psi(\psi)Z(z)$
$R=CJ_p(qr)+DJ_{-p}(qr)$
or* Bessel functions
$R=CJ_n(qr)+DY_n(qr)$
$\Psi=C\sin p\psi+D\cos p\psi$
$Z=C\sin[(k^2-q^2)^{\frac{1}{2}}z]+D\cos[(k^2-q^2)^{\frac{1}{2}}z]$

For $\mathbf{A}=\mathbf{A}(r, z)$, A_r or $A_\psi=R(r)Z(z)$

$R=CJ_1(qr)+DY_1(qr)$
$Z=C\sin[(k^2-q^2)^{\frac{1}{2}}z]+D\cos[(k^2-q^2)^{\frac{1}{2}}z]$

For $\mathbf{A}=\mathbf{A}(r, \psi)$, $A_z=R(r)\Psi(\psi)$

$R=CJ_p(ikr)+DJ_{-p}(ikr)$
or*
$R=CJ_n(ikr)+DY_n(ikr)$
$\Psi=C\sin p\psi+D\cos p\psi$

3. *Elliptic-cylinder coordinates.* $\mathbf{A}=\mathbf{A}(\eta, \theta)$, $A_z=H(\eta)\Theta(\theta)$

$H=DCe_m(\eta, q)+EFe_m(\eta, q)$
or
$H=DSe_m(\eta, q)+EGe_m(\eta, q)$
} Mathieu functions
$\Theta=Dce_m(\theta, q)+Efe_m(\theta, q)$
or
$\Theta=Dse_m(\theta, q)+Ege_m(\theta, q)$

where $q=k^2c^2/4$ and $a=\alpha_2-k^2c^2/2$ is adjusted so that the solution has a period π or 2π.

4. *Parabolic-cylinder coordinates.* $\mathbf{A}=\mathbf{A}(\mu, \nu)$, $A_z=M(\mu)N(\nu)$

$M=ED_p(iq\mu)+FD_{-p-1}(q\mu)$
Weber functions
$N=ED_p(q\nu)+FD_{-p-1}(iq\nu)$

where $k^2=-q^2/4$.

5. *Spherical coordinates.* $\mathbf{A}=\mathbf{A}(r, \theta)$, $A_\psi=R(r)\Theta(\theta)$

$R=(1/r)^{\frac{1}{2}}[CJ_{p+\frac{1}{2}}(kr)+DJ_{-(p+\frac{1}{2})}(kr)]$
$\Theta=CP_p^1(\cos\theta)+DQ_p^1(\cos\theta)$
Legendre functions

6. *Prolate spheroidal coordinates.* $\mathbf{A}=\mathbf{A}(\eta, \theta)$, $A_\psi=H(\eta)\Theta(\theta)$

$H=CP_p^1(kc, \cosh\eta)+DQ_p^1(kc, \cosh\eta)$
} Legendre wave functions
$\Theta=CP_p^1(kc, \cos\theta)=DQ_p^1(kc, \cos\theta)$

7. *Oblate spheroidal coordinates.* $\mathbf{A}=\mathbf{A}(\eta, \theta)$, $A_\psi=H(\eta)\Theta(\theta)$

$H=CP_p^1(ikc, i\sinh\eta)+DQ_p^1(ikc, i\sinh\eta)$
} Legendre wave functions
$\Theta=CP_p^1(ikc, \cos\theta)+DQ_p^1(ikc, \cos\theta)$

8. *Parabolic coordinates.* $\mathbf{A}=\mathbf{A}(\mu, \nu)$, $A_\psi=M(\mu)N(\nu)$

$M=CJ_1(k, iq\mu)+DY_1(k, iq\mu)$
} Bessel wave functions
$N=CJ_1(k, q\nu)+DY_1(k, q\nu)$

* If p is an integer, the second solution must be employed.

The remaining components of $\mathbf{E}_0$ and $\mathbf{H}_0$ are found by making use of the Maxwell equations. Since E_{0z} and H_{0z} contain an exponential function of z, the divergence conditions can be satisfied only if E_{0r}, $E_{0\psi}$, H_{0r}, and $H_{0\psi}$ contain the same functions of z. Substitution of this result into the curl equations gives

$$\begin{cases} \pm(q^2-\Gamma^2)^{\frac{1}{2}}E_{0\psi}-i\omega\mu H_{0r}=\partial E_{0z}/r\partial\psi, \\ -(\sigma+i\omega\epsilon)E_{0\psi}\pm(q^2-\Gamma^2)^{\frac{1}{2}}H_{0r}=\partial H_{0z}/\partial r, \end{cases} \quad (12)$$

$$\begin{cases} \pm(q^2-\Gamma^2)^{\frac{1}{2}}E_{0r}+i\omega\mu H_{0\psi}=\partial E_z/\partial r, \\ (\sigma+i\omega\epsilon)E_{0r}\pm(q^2-\Gamma^2)^{\frac{1}{2}}H_{0\psi}=\partial H_z/r\partial\psi. \end{cases} \quad (13)$$

Equations (12) and (13) are linear equations in the remaining components of the field vectors and permit the completion of the solution.

IV. CYLINDRICAL FIELDS INDEPENDENT OF z

In many cylindrical problems the solution is independent of z. In this case we wish to investigate the possibility of using other coordinate systems. Let $\mathbf{A}=\mathbf{A}(u^1, u^2)$. For a cylindrical coordinate system the metric coefficients can always be written,

$$g_{11}=g_{11}(u^1, u^2)=g_{22}, \quad g_{33}=1, \quad g^{\frac{1}{2}}=(g_{11}g_{22})^{\frac{1}{2}}. \quad (14)$$

In this case the third component of the vector Helmholtz equation becomes

$$[\partial^2A_z/(\partial u^1)^2]+[\partial^2A_z/(\partial u^2)^2]+g_{11}k^2A_z=0. \quad (15)$$

Now let $A_z=U^1(u^1)U^2(u^2)$. Then Eq. (15) becomes

$$[d^2U^1/U^1(du^1)^2]+[d^2U^2/U^2(du^2)^2]+k^2g_{11}=0; \quad (16)$$

but the necessary and sufficient condition for separability of Eq. (16) is

$$g_{11}=h_1(u^1)+h_2(u^2). \quad (17)$$

Theorem III.—If $\mathbf{A}=\mathbf{A}(u^1, u^2)$ in a cylindrical coordinate system, then the vector Helmholtz equation in A_z will separate provided the metric coefficients can be written in the form

$$g_{11}=g_{22}=h_1(u^1)+h_2(u^2).$$

This condition is satisfied in rectangular, circular-cylinder, elliptic-cylinder, and parabolic-cylinder coordinates.

Once the general form of solution for the z components, E_{0z} and H_{0z}, has been found, the remaining components are determined from the Maxwell equations which give, in this case,

$$\begin{cases} E_{01}=\dfrac{1}{(\sigma+i\omega\epsilon)(g_{11})^{\frac{1}{2}}}\dfrac{\partial H_{0z}}{\partial u^2}, \\ E_{02}=-\dfrac{1}{(\sigma+i\omega\epsilon)(g_{11})^{\frac{1}{2}}}\dfrac{\partial H_{0z}}{\partial u^1}, \\ H_{01}=\dfrac{i}{\omega\mu(g_{11})^{\frac{1}{2}}}\dfrac{\partial E_{0z}}{\partial u^2}, \\ H_{02}=-\dfrac{i}{\omega\mu(g_{11})^{\frac{1}{2}}}\dfrac{\partial E_{0z}}{\partial u^1}. \end{cases} \quad (18)$$

V. AXIALLY SYMMETRIC FIELDS

Consider an axially symmetric field in which the solution is independent of the angle ψ, $\mathbf{A}=\mathbf{A}(u^1, u^2)$. In an axially symmetric coordinate system, the metric coefficients can always be written,

$$\begin{aligned} g_{11}&=g_{22}(u^1, u^2),\\ g_{33}&=g_{33}(u^1, u^2). \end{aligned} \tag{19}$$

The ψ-component of the vector Helmholtz equation is now

$$\begin{aligned}&(\partial/\partial u^1)[(1/g_{33})^{\frac{1}{2}}(\partial/\partial u^1)((g_{33})^{\frac{1}{2}}A_\psi)]\\ &+(\partial/\partial u^2)[(1/g_{33})^{\frac{1}{2}}(\partial/\partial u^2)((g_{33})^{\frac{1}{2}}A_\psi)]+k^2g_{11}A_\psi=0. \end{aligned} \tag{20}$$

The assumption of $A_\psi=U^1(u^1)U^2(u^2)$, in Eq. (20) leads to

$$\begin{aligned}&(\partial/U^1\partial u^1)(1/g_{33})^{\frac{1}{2}}[(\partial/\partial u^1)((g_{33})^{\frac{1}{2}}U^1)]\\ &+(\partial/U^2\partial u^2)[(1/g_{33})^{\frac{1}{2}}(\partial/\partial u^2)((g_{33})^{\frac{1}{2}}U^2)]+k^2g_{11}=0. \end{aligned} \tag{21}$$

Necessary conditions for the separation of Eq. (21) are

$$\begin{aligned} (g_{33})^{\frac{1}{2}}&=f_1(u^1)f_2(u^2),\\ g_{11}&=h_1(u^1)+h_2(u^2). \end{aligned} \tag{22}$$

These conditions can also be shown to be sufficient by direct substitution of Eq. (22) into Eq. (21).

Theorem IV.—Necessary and sufficient conditions for the separation of the ψ-component of the vector Helmholtz equation, in rotational coordinates and with a symmetric field $\mathbf{A}=\mathbf{A}(u^1, u^2)$, are

$$\begin{cases} g_{11}=g_{22}=h_1(u^1)+h_2(u^2),\\ (g_{33})^{\frac{1}{2}}=f_1(u^1)f_2(u^2). \end{cases}$$

These conditions are satisfied in spherical, prolate spheroidal, oblate spheroidal, and parabolic coordinate systems.

The remaining components are found in terms of $E_{0\psi}$ and $H_{0\psi}$ from the relations

$$\begin{aligned} E_{01}&=[1/(\sigma+i\omega\epsilon)(g_{11}g_{33})^{\frac{1}{2}}](\partial/\partial u^2)((g_{33})^{\frac{1}{2}}H_{0\psi}),\\ E_{02}&=[-1/(\sigma+i\omega\epsilon)(g_{11}g_{33})^{\frac{1}{2}}](\partial/\partial u^1)((g_{33})^{\frac{1}{2}}H_{0\psi}),\\ H_{01}&=[i/\omega\mu(g_{11}g_{33})^{\frac{1}{2}}](\partial/\partial u^2)((g_{33})^{\frac{1}{2}}E_{0\psi}),\\ H_{02}&=[-i/\omega\mu(g_{11}g_{33})^{\frac{1}{2}}](\partial/\partial u^1)((g_{33})^{\frac{1}{2}}E_{0\psi}). \end{aligned} \tag{23}$$

VI. SUMMARY

Separability conditions have been derived for the vector Helmholtz equation. If the assumption is made that the vector Helmholtz equation will not separate in any coordinate systems in which the scalar Helmholtz equation does not separate, then the following conclusions can be drawn.

I. The only coordinate system in which all of the components of the vector Helmholtz equation separate is rectangular coordinates. Expressions for all components of the vector field can be found directly from the Helmholtz equation.

II. In circular-cylinder coordinates, with

$$\mathbf{A}=\mathbf{A}(u^1, u^2, u^3),$$

the differential equation in A_z is separable. The remaining components of the vector field are found from the Maxwell equations.

III. In cylindrical coordinate systems, separation of the A_z-equation is possible if the vector field is independent of z and if

$$g_{11}=g_{22}=h_1(u^1)+h_2(u^2).$$

This type of separation occurs in elliptic-cylinder and parabolic-cylinder coordinates, in addition to rectangular and circular-cylinder coordinates. The other components of the field are found from Maxwell's equations.

IV. In rotational coordinate systems, separation of the A_ψ equation is possible if the vector field is independent of ψ and if

$$g_{11}=g_{22}=h_1(u^1)+h_2(u^2) \quad \text{and} \quad (g_{33})^{\frac{1}{2}}=f_1(u^1)f_2(u^2).$$

These conditions are satisfied in spherical, prolate and oblate spheroidal, and parabolic coordinates.

Thus, there appear to be eight coordinate systems in which separability methods are applicable to electromagnetic problems. These are shown in Table I together with the solutions of the separated differential equations.

Diffraction Errors in an Optical Measurement at Radio Wavelengths

G. A. Woonton, J. A. Carruthers, H. A. Elliott, and E. Caroline Rigby
Eaton Electronics Research Laboratory, Department of Physics, McGill University, Montreal, Canada
(Received July 12, 1950)

The errors arising in the measurement of antenna patterns by means of a lens are examined. The effect of the lens aperture is separated from the effect of lens aberrations and only the effect of the aperture is discussed. Theoretical solution of the problem is obtained for a uniformly illuminated aperture and for short electromagnetic horns. The theory is found to be in agreement with experimental measurements. The phenomenon of mutilation due to the aperture is examined with the help of both computed results and experimental measurements. The systematic errors due to mutilation have been found and have been shown as precision limits. An extrapolation of the data suggests that a lens forty wavelengths wide can be used for the precision measurement of the patterns of antennas ten to twenty wavelengths wide. The technique is entering its range of usefulness at a wavelength of 3.2 cm and should increase in value as the wavelength is decreased.

INTRODUCTION

MANY techniques which have been brought to perfection for optical measurements are still unexploited at radio wavelengths. It is probable that they have been neglected for lack of exact information on the nature and magnitude of the errors which are associated with them at the longer wavelengths. The need for information on these errors is now becoming more apparent. The measurements and computations which are included in this paper, and a report on a preliminary set of experiments[1] along with a theoretical discussion[2] form part of an investigation of the errors that occur when an optical technique is used to measure antenna patterns.

An optical system for measuring antenna patterns has been described in an earlier paper.[1] The antenna under measurement was rotated in a field which had been collimated by a metal lens. Figures 4 and 5 of that paper led to the conclusion that, if allowance were made for random errors of measurement, the residual errors could be traced to diffraction of the incident wave by the relatively small lens aperture. Whether or not this conclusion was completely valid, it seems clear that systematic errors can arise from only two sources: (1) faulty summation of the incident energy due to imperfections in the lens and (2) diffraction from the aperture of the lens.

This paper is concerned with the second effect only. In all the experiments, the antenna was placed close to a limiting aperture in a metal sheet, and its pattern was measured by a receiver some hundreds of feet beyond.

The theory of measurement through apertures shows that the distant pattern is controlled by seven variables; it is clear that a very great number of experiments would be required to study the effect of changes in the various parameters by sufficiently small increments. For this reason it has been necessary to rely, in part, on computed results. Because the general solution for the mutilated pattern has been obtained for a linear antenna only, both computations and experimental measurements have been confined to linear antennas. Even so, the results can be expected to indicate the effect of the system on the measured pattern of the most general type of antenna.

The general solution of this problem is discussed elsewhere;[2] but two specialized solutions, one for a uniformly illuminated aperture, the other for the *E*-plane of a short horn, are discussed in the appendix to this paper. The solution for the *E*-plane of a short horn applies to radiators that can be realized with precision in experiment, but the solution is so involved that it can be evaluated only by a very time-consuming process.

The solution for the uniform aperture has been obtained in a form which does not require a prohibitive effort for its evaluation. Fortunately, an aperture distribution in which the phase is constant is an ideal that is approximated by many real antennas so that a study of the uniformly illuminated aperture leads to quantitative information that is of practical value. These considerations have led to the following organization of the investigation:

(1) The general theory has been verified by an experimental test of the relation derived for the *E*-plane of a short horn.
(2) The effect of the geometrical parameters on the pattern has been studied for the uniformly illuminated aperture by computation.
(3) The effect of non-uniform antenna illumination has been studied by experimental measurements on short horns.

The discussion in this paper is related to several other problems, although no attempt has been made here to apply the analysis directly to them. The use of curved mirrors for the measurement of antenna patterns, of plane mirrors in scanning problems, and of mirrors or lenses for the measurement of the energy back-scattered by an object in an electromagnetic field are all in this category.

THEORY

The distant field of a radiator, when measured through a limiting aperture, is predicted by the follow-

[1] Woonton, Borts, and Carruthers, J. Appl. Phys. **21**, 428 (1950).
[2] G. A. Woonton, J. Appl. Phys. **21**, 577 (1950).

ing relation: (reference 2)

$$G_{1am}(\gamma)=c\int_{-\infty}^{+\infty}G_1\left(\beta+\frac{\sin\alpha}{\lambda}\right)\times\frac{\sin\pi c(\beta-\gamma)}{\pi c(\beta-\gamma)}\exp[j\pi b\lambda(\beta^2-\gamma^2)]d\beta. \quad (1)$$

The symbols that are used in this equation, and some others that will appear later, are illustrated by Fig. 1 and are defined as follows: A = a constant, a = width of the antenna aperture in the plane of rotation, b = distance from the center of the antenna aperture to the center of the limiting aperture, c = width of the limiting aperture in the plane of rotation of the antenna, $F_1(x)$ = the field distribution across the aperture of the antenna, $G_1[\beta+(\sin\alpha/\lambda)]$ = the unmutilated distant field of the antenna, $G_{1am}(\gamma)$ = the mutilated distant field of the antenna, $j=(-1)^{\frac{1}{2}}$, l = the length of a horn radiator measured from the equivalent vertex to the aperture plane, α = the angle through which the radiator has been rotated, measured in the plane of rotation, β = a dummy symbol used to represent γ, $\gamma=-\sin\theta/\lambda$, and θ = the angle between the line b and the radius vector from the origin to a field point. Methods of integrating Eq. (1) are discussed in the appendix. The end result of each integration is given below.

Case 1—The Uniformly Illuminated Aperture

The aperture illumination function is given by

$$F_1(x)=A \quad \text{for} \quad |x|<(\tfrac{1}{2}a)$$
$$=0 \quad \text{for} \quad |x|>(\tfrac{1}{2}a).$$

Equation (1) then integrates to

$$G_{1am}(0)=\lambda\left(\frac{j}{2}\right)^{\frac{1}{2}}\left(A\frac{a}{\lambda}\right)\frac{\sin[\pi(a/\lambda)\sin\alpha]}{\pi(a/\lambda)\sin\alpha}\times\{[C(R)+C(T)]-j[S(R)+S(T)]\}$$
$$-j\frac{A\lambda}{2\pi\sin\alpha}\left(\frac{j}{2}\right)^{\frac{1}{2}}\exp\left(j\pi\frac{b}{\lambda}\sin^2\alpha\right)\times\left[\exp\left(-j\pi\frac{c}{\lambda}\sin\alpha\right)\{[C(P-R)-C(P-T)]-j[S(P-R)-S(P-T)]\}\right.$$
$$\left.-\exp\left(j\pi\frac{c}{\lambda}\sin\alpha\right)\{[C(P+T)-C(P+R)]-j[S(P+T)-S(P+R)]\}\right]. \quad (2)$$

The following new symbols have been introduced into Eq. (2): C = Fresnel's cosine integral, S = Fresnel's sine integral, $P=(b/\lambda)(2\lambda/b)^{\frac{1}{2}}\sin\alpha$, $R=\frac{1}{2}(2\lambda/b)[^{\frac{1}{2}}(c/\lambda)+(a/\lambda)]$, and $T=\frac{1}{2}(2\lambda/b)^{\frac{1}{2}}[(c/\lambda)-(a/\lambda)]$.

Case 2—The *E*-Plane of a Short, Horn Radiator

The aperture illumination function is given by

$$F_1(x)=F_A(x)\exp(-j\pi x^2/\lambda l)$$

where

$$F_A=A \quad \text{for} \quad |x|\leqslant\tfrac{1}{2}a$$
$$=0 \quad \text{for} \quad |x|>\tfrac{1}{2}a.$$

Equation (1) can then be integrated into the approximate form:

$$G_{1am}(0)=G_1\left(\frac{\sin\alpha}{\lambda}\right)+\frac{A\lambda}{2\pi}(j-1)\left(\frac{b}{2\lambda}\right)^{\frac{1}{2}}D, \quad (3)$$

where

$$D=\sum_{i=1}^{2}\frac{1}{1+j(K/2\pi L_i)}\left\{-\frac{K}{L_i}\int_{\frac{1}{2}(c/\lambda-a/\lambda)}^{\frac{1}{2}(c/\lambda+a/\lambda)}\exp[-j\pi(KZ^2+2L_iZ+M_i)]dZ+\left[\frac{\exp[-j\pi(KZ^2+2L_iZ+M_i)]}{2\pi L_iZ}\left(j-\frac{1}{2\pi L_iZ}\right)\right]_{\frac{1}{2}[(c/\lambda)-(a/\lambda)]}^{\frac{1}{2}[(c/\lambda)+(a/\lambda)]}\right\}$$

and

$$K=\lambda/b+\lambda/l$$

$$L_1=-\frac{1}{2}\frac{\lambda}{l}\frac{c}{\lambda}+\sin\alpha \qquad L_2=-\frac{1}{2}\frac{\lambda}{l}\frac{c}{\lambda}-\sin\alpha$$

$$M_1=-\frac{1}{4}\frac{\lambda}{l}\left(\frac{c}{\lambda}\right)^2-\frac{c}{\lambda}\sin\alpha \qquad M_2=-\frac{1}{4}\frac{\lambda}{l}\left(\frac{c}{\lambda}\right)^2+\frac{c}{\lambda}\sin\alpha.$$

Equations (2) and (3) are both universal to the extent that linear dimensions are measured in wavelengths.

Figure 2 presents a comparison of theory and measurement when the patterns of two of the horns were mutilated by the larger aperture. At any angle between zero and twenty degrees it is estimated that the probable error in the power measurement is 0.25 db. The

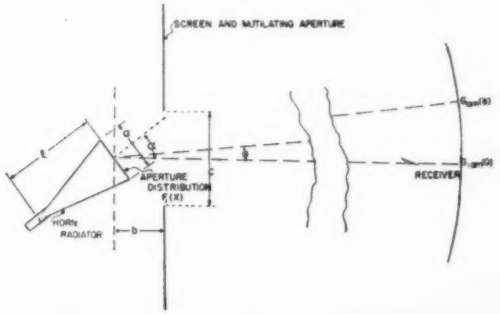


FIG. 1. A plan view illustrating the parameters of the system.

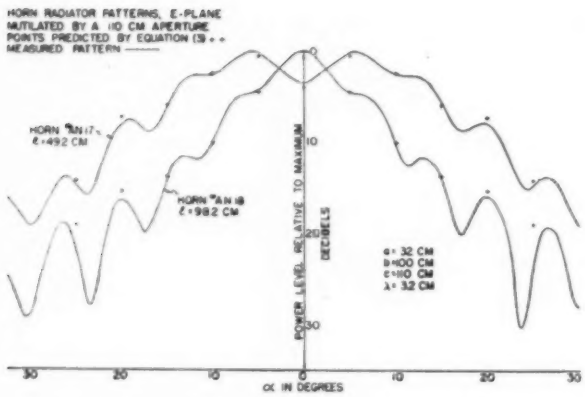


FIG. 2. Comparison of predicted and measured points. $c=110$ cm; two values of l; $a=32$ cm.

probable error, calculated from the deviations between the computed and measured points at 0, ± 5, ± 10, ± 15, and ± 20 degrees, was found to be 0.34 db for the 49.2-cm horn and 0.28 db for the 98.2-cm horn. These results suggest that disagreements between theory and measurement are due to random experimental errors.

Figure 3 presents a comparison between theory and measurement when the patterns of the same two horns were mutilated by the smaller aperture. The probable error calculated from the deviations between computed and measured points was found to be 0.41 db for the 49.2-cm horn, and 0.74 db for the 98.2-cm horn.

The data displayed by Fig. 4 were compiled from pattern measurements made through the same mutilating apertures but on a horn of only half the previous width. The probable error in the computed points was found to be 0.58 db for the larger mutilating aperture and 0.33 db for the smaller.

It has been concluded from the data presented in these graphs that patterns predicted by Eq. (3) are as accurate as the experimental measurements when the larger mutilating aperture is considered, and not much less accurate for the smaller.

The physical characteristics of the mutilating apertures and the experiments, from which the data in Figs. 2–4 were obtained, are described in the section on experimental results.

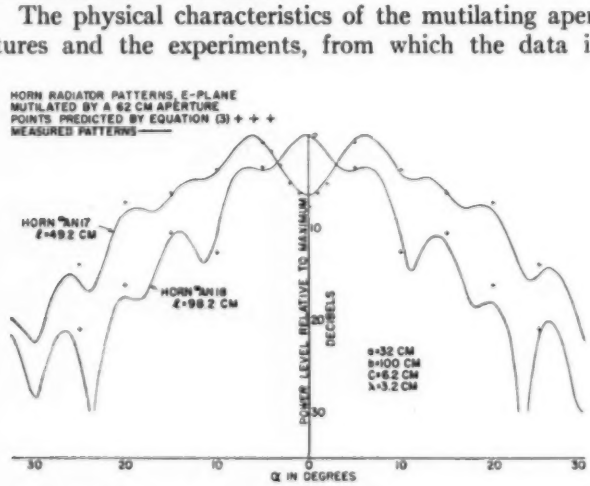


FIG. 3. Comparison of predicted and measured points. $c=62$ cm; $a=32$ cm; two values of l.

COMPUTED RESULTS

The data which are presented and discussed in this section have all been computed from Eq. (2). They relate to the manner in which a mutilating screen modifies the distant pattern of a uniformly illuminated, radiating aperture. Because Eq. (2) was derived from Eq. (1) in closed form, predictions made from it should be as valid as Eq. (1) and more accurate than the predictions made from Eq. (3), which were verified by experiment.

Figure 5 is a contour diagram which relates the power pattern of the distant field to the width of the antenna when the pattern is measured through a limiting aperture of fixed width placed in a fixed position in front of the antenna. The contours are lines of constant power measured in decibels below the value of maximum power. The abscissas are in units of antenna aperture

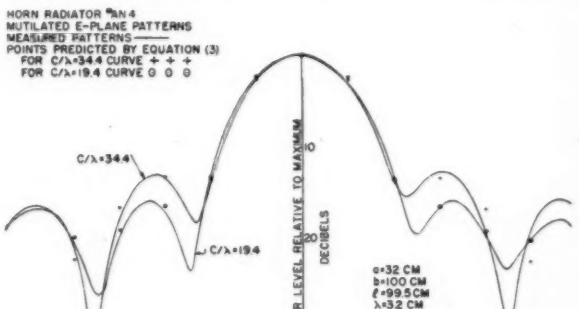


FIG. 4. Comparison of predicted and measured points. $a=16$ cm; $l=99.5$ cm; two values of c.

width, a/λ. Equation (2) shows that, in the absence of mutilation, the relative power patterns are independent of a/λ when they are expressed as functions of the variable $(\pi a/\lambda \sin\alpha)$. The ordinates of the contour diagram have been expressed in units of this variable. Plotted in this way, any region of the diagram in which the contours are straight and parallel to the axis of a/λ is a region in which the limiting aperture has no effect. Elsewhere, the deviations of the contours from the straight-line relation are a measure of the mutilation which has been caused by the limiting aperture. Any cut through the diagram along a line of constant a/λ gives the mutilated distant pattern appropriate to the various parameters.

The following procedure was adopted for estimating the systematic deviation of a mutilated pattern from the true, unmutilated pattern: the true pattern, and the mutilated patterns for the various values of a/λ were plotted in coordinates of relative *db* against $(\pi a/\lambda \sin\alpha)$. The choice of scale was such that one decibel was spread over d linear units of the graph paper, and one unit of

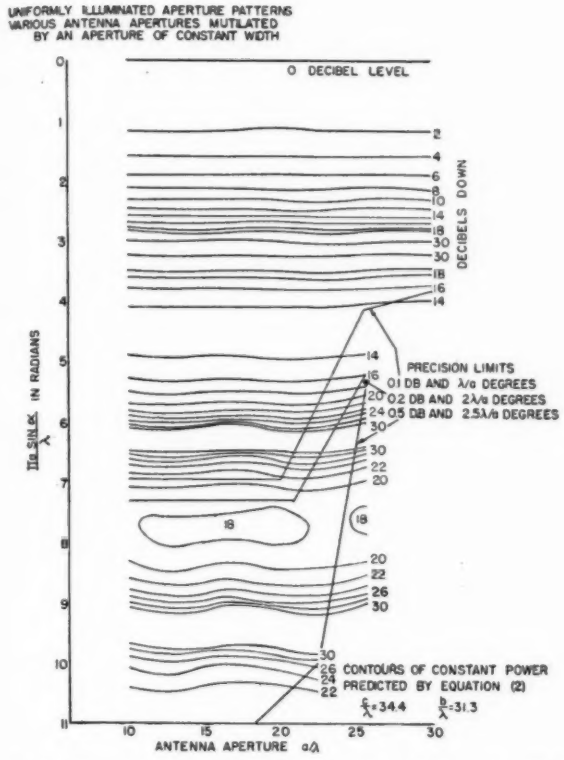


FIG. 5. Computed patterns $l/\lambda = \infty$; a/λ variable; $c/\lambda = 34.4$.

the angular function over nd linear units. A circle of p linear units was drawn around each point of the true pattern. The true pattern and a mutilated pattern were then superimposed with the zero-decibel points coinciding. The mutilated pattern is said to be exact to within p/nd radians and p/d decibels over the region in which it falls everywhere on or within the circumference of these circles. Since the curves are nearly vertical everywhere except at the tops of the side lobes, one can conclude that in most regions this statement of precision is equivalent to the statements: at a given decibel level, the angular position of the mutilated curve is within p/nd radians of its true position; at the tops of the side lobes the curve is within p/d decibels of its true level.

Precision limits found by this procedure were plotted on Fig. 5. Any combination of parameters which can be represented by a point in the area above the limit will give rise to a point in the mutilated pattern which is at least as precise as the values shown on the limit. The limits have been drawn as straight lines and rule off the region where points of lower precision will be encountered; because they do not follow the oscillations of precision, many points can be found below the limits for which the precision is as high as in the area above.

Figure 6 is a contour diagram which relates the power pattern of the distant field of an antenna of fixed width to the width of the limiting aperture through which it is measured. The contours are again lines of constant power. The abscissas are in units of limiting aperture width, c/λ. Unlike Fig. 5, the ordinates are given directly in degrees; but because of the different variable, c/λ, the interpretation of the diagram is the same as for Fig. 5. A cut through the diagram along any line of constant c/λ gives the mutilated distant pattern for that aperture width.

Precision limit lines have been plotted on Fig. 6 and are to be interpreted in the same way as before except that a constant rather than a percentage angular precision is given for all regions of the diagram. The limit lines of Fig. 6 fail to exclude two exceptional points of lower precision; the limits as drawn give a truer picture of the behavior of the system as a whole.

EXPERIMENTAL RESULTS

The equipment used to make the measurements was very little different from that used in many laboratories for the measurement of antenna patterns. The fixed installations at the measuring site consisted of a turntable mounted about thirty feet above the ground on the roof of a hut, and a wooden tower set in a clear space at a distance of two hundred and fifty feet from the turntable. The antenna being measured was mounted on the turntable and was used as the transmitter. Its field was measured by a receiving antenna mounted on the tower in horizontal line with the transmitter.

The addition of a mutilating aperture between the transmitting and receiving antennas constituted the

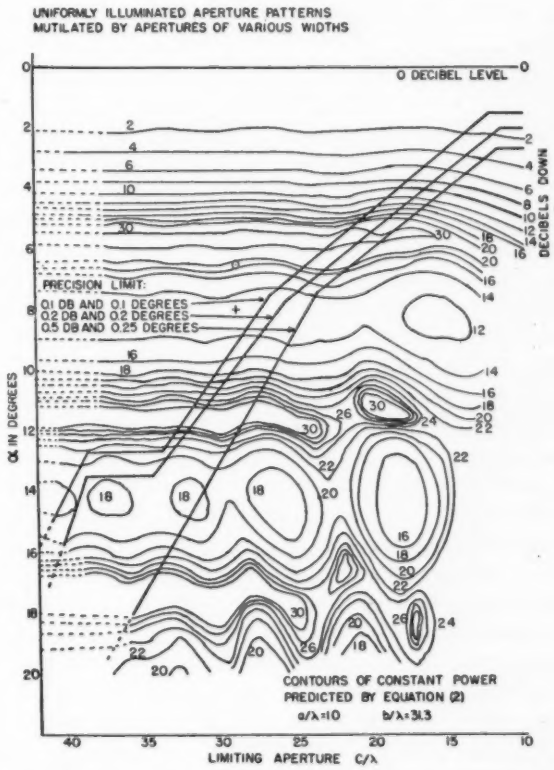


FIG. 6. Computed patterns $l/\lambda = \infty$; $a/\lambda = 10$; c/λ variable. The point marked ⊙ is exceptional in that the precision at this point is less than 0.1 db and 0.1 degree; in the same way the precision is less than 0.2 db and 0.2 degree at the point marked +.

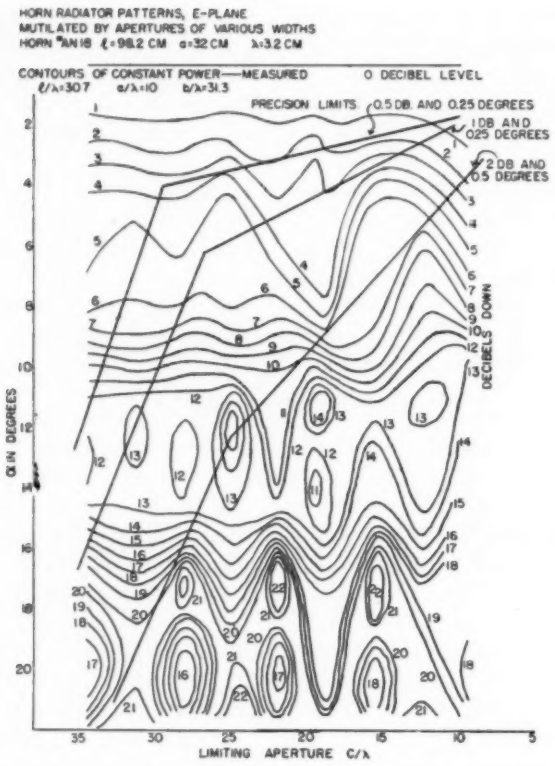

HORN RADIATOR PATTERNS, E-PLANE
MUTILATED BY APERTURES OF VARIOUS WIDTHS
HORN "AN 18 ℓ=98.2 CM a=32 CM λ=3.2 CM
CONTOURS OF CONSTANT POWER—MEASURED
ℓ/λ=30.7 a/λ=10 b/λ=31.3
0 DECIBEL LEVEL
PRECISION LIMITS 0.5 DB. AND 0.25 DEGREES
1 DB AND 0.25 DEGREES
2 DB AND 0.5 DEGREES
DECIBELS DOWN
α IN DEGREES
LIMITING APERTURE C/λ

FIG. 7. Measured patterns $l/\lambda = 30.7$; $a/\lambda = 10$; c/λ variable.

sole difference between this setup and that used to make antenna pattern measurements. A large, flat frame, about ten feet square, was erected at a distance of 100 cm in front of the center of rotation of the turntable. The frame was set with its plane at right angles to the line joining the centers of the transmitting and receiving antennas and was centered both vertically and horizontally on this line. Except for the center region, which was left unobstructed, the frame was covered with copper screen, which prevented stray radiation from reaching the receiver. The mutilating apertures were hung over the center opening in the frame.

All measurements were made at a wavelength of 3.2 cm. Power was supplied to the various antennas by a type 723 A/B klystron which was modulated at 600 cps by a square-wave generator. The power output from the klystron was made independent of reaction from the mutilating screen by means of a 5-db pad placed between the klystron and the horn. The receiving antenna was a 30″ copper paraboloid, with a double-dipole at its focus. The power received by the paraboloid was demodulated by a bolometer or, in some of the experiments, by a crystal. The audio signal was then passed through a narrow band pre-amplifier in the tower, was returned to the location of the turntable by land line, and was then amplified to suitable level for recording. Some of the data were recorded from meter readings, by hand; the rest were automatically recorded by a system which has been described by Carruthers.[3]

[3] J. A. Carruthers, Can. J. Research **28A**, 287 (1950).

The probable error in the relative power measurements has been estimated to be 0.25 db. The probable error in the angular measurements has been estimated to be not worse than 0.25 degree and probably much better. The turntable can be set and read to an angular accuracy of ± 0.05 degree; but, because of imprecisions in reading the recorder patterns, the angular precision of the data is much less than this value. Experimental errors, which were not present in the unmutilated patterns, appeared when a mutilating aperture was used. Detectable irregularities and asymmetries were present in the mutilated pattern when the mutilating aperture was as little as one-half centimeter out of alignment.

Figures 7, 8, and 9 of this section and Figs. 5 and 6 of the preceding section should be considered together. Although the data for Fig. 6 were computed, and those for Figs. 7–9 were measured, all the figures have been compiled by the same method and can be interpreted in the manner described in connection with Fig. 6. Considered together, the five graphs give information on the systematic errors due to mutilation as a function of the angular displacement of the horn, the width of the horn, the non-uniformity of the phase distribution across its mouth, and the width of the mutilating aperture. From the data in these graphs, it has been concluded that, with c/λ fixed, the effects of mutilation on the pattern are stationary over a range of the variable a/λ. The extent of mutilation is not very dependent on the phase distribution. If the measurement is made through a

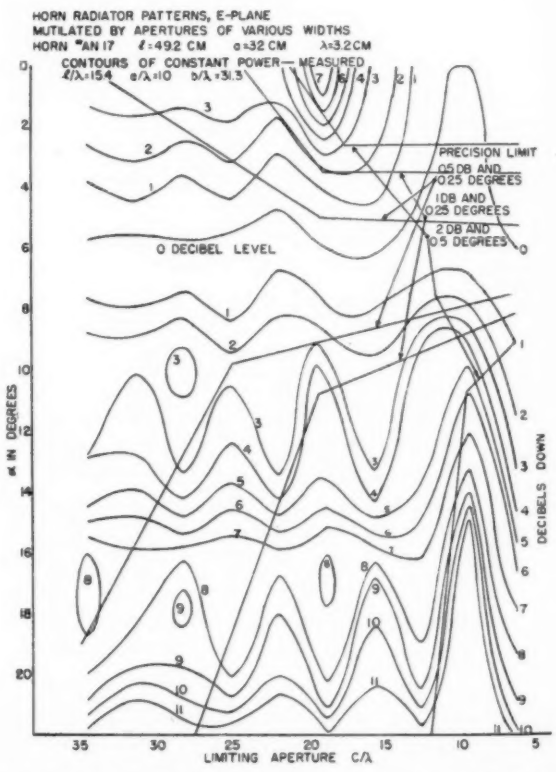

HORN RADIATOR PATTERNS, E-PLANE
MUTILATED BY APERTURES OF VARIOUS WIDTHS
HORN "AN 17 ℓ=49.2 CM a=32 CM λ=3.2 CM
CONTOURS OF CONSTANT POWER—MEASURED
ℓ/λ=15.4 a/λ=10 b/λ=31.3
PRECISION LIMIT
0.5 DB AND 0.25 DEGREES
1 DB AND 0.25 DEGREES
2 DB AND 0.5 DEGREES
0 DECIBEL LEVEL
α IN DEGREES
DECIBELS DOWN
LIMITING APERTURE C/λ

FIG. 8. Measured patterns $l/\lambda = 15.4$; $a/\lambda = 10$; c/λ variable.

mutilating aperture 34λ wide, and if an error of 1 db at the peaks of the side lobes can be tolerated, then patterns of antennas 10λ to 20λ wide can be measured with precision over an angular interval that includes the first two side lobes. If the mutilating aperture is 40λ wide, extrapolation of the data suggests that a precision of 0.5 db and 0.25 degree is possible over the same angular interval for the same range of antenna widths.

Figures 10 and 11 have been plotted in the same coordinates. The abscissas are expressed in units of $(\pi a/\lambda \sin\alpha)$. On this scale, the minima of the radiation patterns occur at the integral multiples of π, which have been marked on each of the component graphs. Although an abscissa of 2π implies an angle of about 24 degrees when $a/\lambda=5$ and an angle of only 6 degrees when $a/\lambda=20$, it represents corresponding points in the two patterns. The ordinates represent the systematic error in decibels caused by the introduction of a mutilating aperture between the antenna and the distant receiver. A positive difference indicates that the relative power at that point in the mutilated pattern is greater than that for the unmutilated pattern.

Figures 10 and 11 show the effect of mutilation on the *E*- and *H*-planes, respectively, of each of the double family of horns when $c/\lambda=34.4$. Similar graphs are available for $c/\lambda=19.4$, but these are of less practical interest because of the heavy mutilation caused by a smaller aperture. The data summarized in Figs. 10 and 11, and in other unpublished graphs, have led to the conclusion that the effects in the *H*-plane are not significantly different from those in the *E*-plane. They have helped to suggest that the phase distribution across the antenna aperture has little effect on the errors due to mutilation. The large error near an abscissa of 2π for horn *AN* 4 illustrates the errors that can be introduced into the measurement of antennas of narrow aperture.

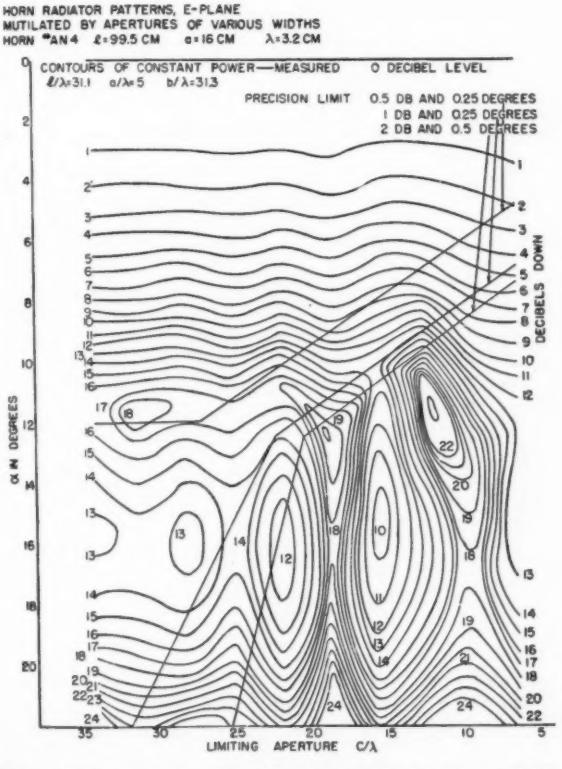

HORN RADIATOR PATTERNS, E-PLANE
MUTILATED BY APERTURES OF VARIOUS WIDTHS
HORN AN 4 ℓ=99.5 CM a=16 CM λ=3.2 CM
CONTOURS OF CONSTANT POWER—MEASURED 0 DECIBEL LEVEL
ℓ/λ=31.1 a/λ=5 b/λ=31.3
PRECISION LIMIT 0.5 DB AND 0.25 DEGREES
1 DB AND 0.25 DEGREES
2 DB AND 0.5 DEGREES
DECIBELS DOWN
α IN DEGREES
LIMITING APERTURE C/λ

FIG. 9. Measured patterns $l/\lambda=31.1$; $a/\lambda=5$; c/λ variable.

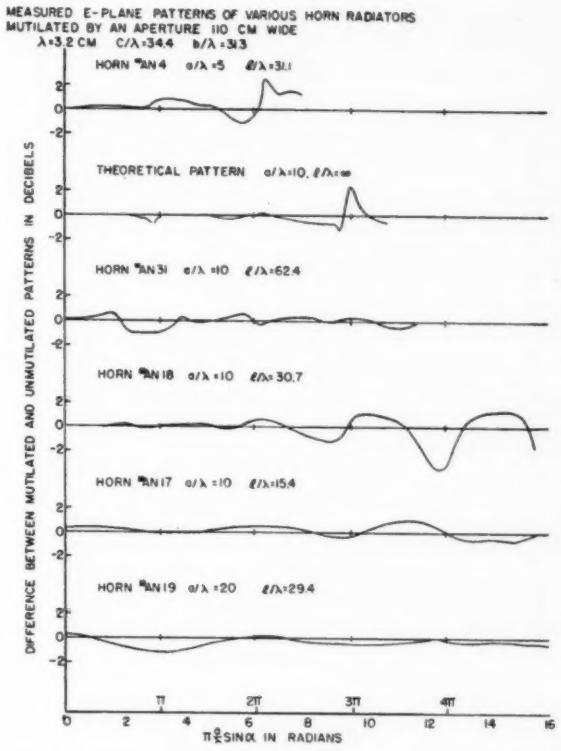

MEASURED E-PLANE PATTERNS OF VARIOUS HORN RADIATORS
MUTILATED BY AN APERTURE 110 CM WIDE
λ=3.2 CM C/λ=34.4 b/λ=33
HORN AN 4 a/λ=5 ℓ/λ=31.1
THEORETICAL PATTERN a/λ=10, ℓ/λ=∞
HORN AN 31 a/λ=10 ℓ/λ=62.4
HORN AN 18 a/λ=10 ℓ/λ=30.7
HORN AN 17 a/λ=10 ℓ/λ=15.4
HORN AN 19 a/λ=20 ℓ/λ=29.4
DIFFERENCE BETWEEN MUTILATED AND UNMUTILATED PATTERNS IN DECIBELS
π SIN α IN RADIANS

FIG. 10. Measured differences due to mutilation, *E*-plane, various l/λ's; various a/λ's; $c/\lambda=34.4$.

Unpublished data shows that a large percentage change in the distance between the antenna and the mutilating screen does not cause a very marked change in the mutilation errors.

CONCLUSIONS

(1) The computations and measurements which have been reported in this paper, when considered in conjunction with the measurements reported in reference (1), lead to the following predictions: an aplanatic lens, forty wavelengths in diameter, may be used for the precision measurement of the pattern of any radiating aperture of dimensions not less than ten and not more than twenty wavelengths over an angular interval that includes the second side lobe. Because the limits on the dimensions of the antenna are set by the ratio of the diameter of the lens to the wavelength, it appears that this technique is entering its range of usefulness at a wavelength of 3.2 cm and should increase in value as the wavelength is decreased.

(2) The predicted effects of mutilation are in good agreement with measurement.

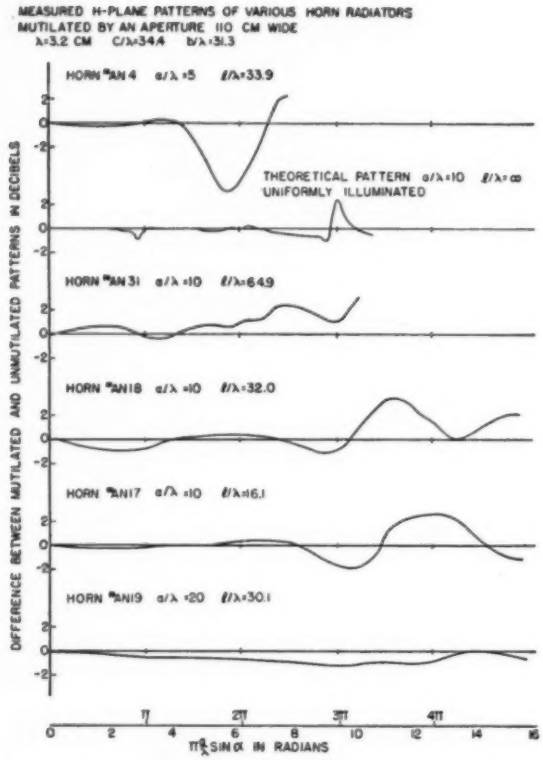

MEASURED H-PLANE PATTERNS OF VARIOUS HORN RADIATORS
MUTILATED BY AN APERTURE 110 CM WIDE
λ=3.2 CM c/λ=34.4 b/λ=31.3
HORN "AN 4 a/λ =5 ℓ/λ=33.9
THEORETICAL PATTERN a/λ=10 ℓ/λ=∞
UNIFORMLY ILLUMINATED
HORN "AN 31 a/λ =10 ℓ/λ=64.9
HORN "AN 18 a/λ =10 ℓ/λ=32.0
HORN "AN 17 a/λ =10 ℓ/λ=16.1
HORN "AN 19 a/λ =20 ℓ/λ=30.1
DIFFERENCE BETWEEN MUTILATED AND UNMUTILATED PATTERNS IN DECIBELS
π/λ SIN α IN RADIANS

FIG. 11. Measured differences caused by mutilation, H-plane, various l/λ's; various a/λ's; $c/\lambda=34.4$.

(3) The extent to which a radiation pattern is mutilated is determined by an involved function of the angular position of the antenna, the dimensions of the antenna, and the dimensions of the mutilating aperture. If patterns are compared in terms of maxima and minima, then with c/λ fixed, the effects of mutilation are stationary over a range of antenna aperture widths. Provided that the pattern contains no central minimum, the modification of the pattern appears first in the outer side lobes and progresses toward the central peak as the parameters of the system are adjusted to make the mutilation more severe. If the pattern contains a central minimum, modification of this region is an early effect of mutilation.

(4) Except for experiments in which the width of the mutilating aperture was about the same as the width of the antenna aperture, each radiation pattern retained its characteristic features, even in the presence of severe mutilation. As a major effect, mutilation modified the relative power at the maxima and minima of the patterns; as a minor effect, it sometimes shifted the angular position of the side lobes through a small angle. For a fixed antenna aperture, the relative power at the maxima and minima of the pattern oscillated around its true value as the width of the mutilating aperture was decreased.

(5) With fixed geometrical parameters, the systematic errors due to mutilation were found to increase as the phase distribution across the antenna aperture departed from a constant, but this was not a major effect.

(6) When measurements were made on the H-plane of horn radiators, it was found that the systematic errors due to mutilation were not significantly different from the errors which appeared in the E-plane measurements. It is concluded that the precision of measurement that is possible with a lens measuring system is not a function of the amplitude distribution across the antenna aperture.

(7) With all other parameters fixed, the changes in the systematic errors were not marked when the distance between the lens and the antenna was increased. It is concluded that the precision of measurement that is possible with a lens measuring system is a slowly-varying function of the distance between the antenna and the lens.

ACKNOWLEDGMENT

This research has been supported in part by the Defence Research Board, Ottawa, Canada, and in part by the United States Air Force through its Cambridge Research Laboratories. The authors wish to acknowledge the help received from Mr. E. Guy Ross in the construction of apparatus, and in the preparation of graphs and diagrams.

APPENDIX.

DERIVATION OF EQUATIONS (2) AND (3)

The general integral is

$$G_{1am}(\gamma)=c\int_{-\infty}^{+\infty} G_1\left(\beta+\frac{\sin\alpha}{\lambda}\right)\frac{\sin\pi c(\beta-\gamma)}{\pi c(\beta-\gamma)} \times\exp[j\pi b\lambda(\beta^2-\gamma^2)]d\beta. \quad (1)$$

The symbols are listed in the section on theory. For all cases considered, $\gamma=0$. Equation (1) may be written

$$G_{1am}(0)=\int_{-\infty}^{+\infty} G_1\left(\beta+\frac{\sin\alpha}{\lambda}\right)\frac{\exp(j\pi b\lambda\beta^2+j\pi c\beta)}{2j\pi\beta}d\beta - \int_{-\infty}^{+\infty} G_1\left(\beta+\frac{\sin\alpha}{\lambda}\right)\frac{\exp(j\pi b\lambda\beta^2-j\pi c\beta)}{2j\pi\beta}d\beta=I(c)-I(-c).$$

In evaluating $I(c)$, it is regarded as a fourier transform in $\exp[j2\pi(c/2)\beta]$. Using pairs 206, 210, 710.0, and 202 as listed by Campbell and Foster,[4] $I(c)$ transforms to

$$I(c)=\left(\frac{j}{b\lambda}\right)^{\frac{1}{2}}\exp\left(-j\pi c\frac{\sin\alpha}{\lambda}\right)\int_{-\infty}^{+\infty}\left[\int_{-\infty}^{z}\exp\left(\frac{-j\pi z^2}{b\lambda}\right)dz\right] \times F_1\left(\frac{c}{2}-x\right)\exp\left(j2\pi x\frac{\sin\alpha}{\lambda}\right)dx,$$

where $F_1(x)\leftrightarrow G_1(\beta)$. The aperture distribution is assumed to be

$$F_1(x)=F_A(x)\phi(x),$$

where

$$F_A(x)=A \quad \text{for} \quad |x|\leqslant\tfrac{1}{2}a$$
$$=0 \quad \text{for} \quad |x|>\tfrac{1}{2}a$$

and

$\phi(x)$ is a phase factor.

---

[4] G. A. Campbell and R. M. Foster, *Fourier Integrals for Practical Applications* (Bell Telephone System Technical Publications, 1942), Monograph B584.

Then

$$I(c)=A\left(\frac{j}{b\lambda}\right)^{\frac{1}{2}}\exp\left(-j\pi c\frac{\sin\alpha}{\lambda}\right)\int_{c/2-a/2}^{c/2+a/2}\left[\int_{-\infty}^{z}\exp\left(\frac{-j\pi z^2}{b\lambda}\right)dz\right]\times\phi\left(\frac{c}{2}-x\right)\exp\left(j2\pi x\frac{\sin\alpha}{\lambda}\right)dx. \quad (4)$$

Case 1—The Uniformly Illuminated Aperture

For this case $\phi(x)=1$. Substituting this into Eq. (4) and using pair 1306.

$$I(c)=A\left(\frac{j}{b\lambda}\right)^{\frac{1}{2}}\frac{\exp[-j\pi c(\sin\alpha/\lambda)]}{j2\pi(\sin\alpha/\lambda)}\times\left\{\exp\left(j2\pi x\frac{\sin\alpha}{\lambda}\right)\int_{-\infty}^{z}\exp\left(\frac{-j\pi z^2}{b\lambda}\right)dz\right]_{c/2-a/2}^{c/2+a/2}-\int_{c/2-a/2}^{c/2+a/2}\exp\left(\frac{-j\pi x^2}{b\lambda}+j2\pi x\frac{\sin\alpha}{\lambda}\right)dx\Bigg\}$$

$$=\frac{A\lambda}{j2\pi\sin\alpha}\left(\frac{j}{2}\right)^{\frac{1}{2}}\left[\exp\left(j\pi a\frac{\sin\alpha}{\lambda}\right)\left\{\frac{1-j}{2}+C(R)-jS(R)\right\}-\exp\left(-j\pi a\frac{\sin\alpha}{\lambda}\right)\left\{\frac{1-j}{2}+C(T)-jS(T)\right\}+\exp\left(j\pi b\lambda\frac{\sin^2\alpha}{\lambda^2}-j\pi c\frac{\sin\alpha}{\lambda}\right)\times\{[C(P-R)-C(P-T)]-j[S(P-R)-S(P-T)]\}\right].$$

$I(-c)$ is found by changing the sign of c throughout this expression. Equation (2) of the section on theory gives the mutilated pattern as the difference $\{I(c)-I(-c)\}$.

Case 2—The *E*-Plane of a Short, Horn Radiator

For this case $\phi(x)=\exp(-j\pi x^2/\lambda l)$. Substituting this into Eq. (4) and integrating with respect to z

$$I(c)=A\left(\frac{j}{2}\right)^{\frac{1}{2}}\exp\left(-j\pi c\frac{\sin\alpha}{\lambda}\right)\int_{c/2-a/2}^{c/2+a/2}\left\{\frac{1-j}{2}+C\left[\left(\frac{2}{b\lambda}\right)^{\frac{1}{2}}x\right]-jS\left[\left(\frac{2}{b\lambda}\right)^{\frac{1}{2}}x\right]\right\}\exp\left[-j\pi\left\{\left(\frac{c}{2}-x\right)^2\frac{1}{\lambda l}-2x\frac{\sin\alpha}{\lambda}\right\}\right]dx$$

$C\left[\left(\frac{2}{b\lambda}\right)^{\frac{1}{2}}x\right]-jS\left[\left(\frac{2}{b\lambda}\right)^{\frac{1}{2}}x\right]$ is given by the asymptotic expansions

$$\frac{1-j}{2}+\left(\frac{b\lambda}{2}\right)^{\frac{1}{2}}\frac{1}{\pi x}\exp\left(\frac{-j\pi x^2}{b\lambda}\right)\left[j-\frac{b\lambda}{2\pi x^2}+\cdots\right]\quad\text{for}\quad x>0$$

or

$$\frac{-(1-j)}{2}+\left(\frac{b\lambda}{2}\right)^{\frac{1}{2}}\frac{1}{\pi x}\exp\left(\frac{-j\pi x^2}{b\lambda}\right)\left[j-\frac{b\lambda}{2\pi x^2}+\cdots\right]\quad\text{for}\quad x<0.$$

Only apertures for which $c>a$ have been considered and, therefore, the asymptotic series for $x>0$ is used in evaluating $I(c)$. On substituting the first two terms of this series, $I(c)$ becomes

$$I(c)=A\int_{-a/2}^{+a/2}\exp\left(\frac{-j\pi u^2}{\lambda l}-j2\pi u\frac{\sin\alpha}{\lambda}\right)du+\frac{A\lambda(j-1)}{2\pi}\left(\frac{b}{2\lambda}\right)^{\frac{1}{2}}\int_{\frac{1}{2}[(c/\lambda)-(a/\lambda)]}^{\frac{1}{2}[(c/\lambda)+(a/\lambda)]}\frac{\exp[-j\pi(Kz^2+2L_2z+M_2)]}{z}dz.$$

The first of these two terms is the unmutilated pattern of the antenna. $I(-c)$ is evaluated by substituting the series appropriate to $x<0$ for the expression $C[(2/b\lambda)^{\frac{1}{2}}x]-jS[(2/b\lambda)^{\frac{1}{2}}x]$. Then

$$I(-c)=\frac{-A\lambda(j-1)}{2\pi}\left(\frac{b}{2\lambda}\right)^{\frac{1}{2}}\int_{\frac{1}{2}[(c/\lambda)-(a/\lambda)]}^{\frac{1}{2}[(c/\lambda)+(a/\lambda)]}\frac{\exp[-j\pi(Kz^2+2L_1z+M_1)]}{z}dz.$$

TABLE I.

	$C[(2/b\lambda)^{\frac{1}{2}}x]-jS[(2/b\lambda)^{\frac{1}{2}}x]$			
	$x=(c+a)/2$		$x=(c-a)/2$	
	True value	Approximation	True value	Approximation
$c=110,\ a=16$	$0.562-j0.482$	$0.561-j0.481$	$0.526-j0.580$	$0.525-j0.581$
$c=62,\ a=16$	$0.575-j0.573$	$0.572-j0.574$	$0.341-j0.435$	$0.345-j0.419$
$c=110,\ a=32$	$0.461-j0.459$	$0.460-j0.460$	$0.575-j0.573$	$0.572-j0.574$
$c=62,\ a=32$	$0.526-j0.580$	$0.525-j0.581$	$0.724-j0.612$	$0.716-j0.660$

Hence,

$$G_{1am}(0)=I(c)-I(-c)=G_1\left(\frac{\sin\alpha}{\lambda}\right)+\frac{A\lambda(j-1)}{2\pi}\left(\frac{b}{2\lambda}\right)^{\frac{1}{2}}\int_{\frac{1}{2}[(c/\lambda)-(a/\lambda)]}^{\frac{1}{2}[(c/\lambda)+(a/\lambda)]}\frac{\exp[-j\pi(Kz^2+2L_1z+M_1)]+\exp[-j\pi(Kz^2+2L_2z+M_2)]}{z}dz. \quad (5)$$

An approximate evaluation of the integral in Eq. (5) is obtained as follows. The general expression

$$E_p(K,L,M)=\int_{\frac{1}{2}[(c/\lambda)-(a/\lambda)]}^{\frac{1}{2}[(c/\lambda)+(a/\lambda)]}\frac{\exp[-j\pi(Kz^2+2Lz+M)]}{z^p}dz \qquad p=0,1,2,\cdots,$$

when integrated by parts leads to

$$E_1(K,L,M)=\left[j\frac{\exp[-j\pi(Kz^2+2Lz+M)]}{2\pi Lz}\right]_{\frac{1}{2}[(c/\lambda)-(a/\lambda)]}^{\frac{1}{2}[(c/\lambda)+(a/\lambda)]}-\frac{KE_0}{L}+\frac{jE_2}{2\pi L} \quad (6)$$

and

$$E_2(K,L,M)=\left[j\frac{\exp[-j\pi(Kz^2+2Lz+M)]}{2\pi Lz}\right]_{\frac{1}{2}[(c/\lambda)-(a/\lambda)]}^{\frac{1}{2}[(c/\lambda)+(a/\lambda)]}-\frac{KE_1}{L}+\frac{jE_3}{\pi L}. \quad (7)$$

Eliminating E_2 between (6) and (7) and solving for E_1 gives

$$E_1(K,L,M)=\frac{1}{[1+(jK/2\pi L^2)]}\left\{\frac{-K}{L}\int_{\frac{1}{2}[(c/\lambda)-(a/\lambda)]}^{\frac{1}{2}[(c/\lambda)+(a/\lambda)]}\exp[-j\pi(Kz^2+2Lz+M)]dz+\left[\frac{\exp[-j\pi(Kz^2+2Lz+M)]}{2\pi Lz}\left(j-\frac{1}{2\pi Lz}\right)\right]_{\frac{1}{2}[(c/\lambda)-(a/\lambda)]}^{\frac{1}{2}[(c/\lambda)+(a/\lambda)]}-\frac{1}{2\pi^2L^2}E_3\right\}.$$

If $(c-a)/2\lambda$ is sufficiently large, the term involving E_3 may be neglected. Omitting this term and substituting the resulting expression for E_1 into Eq. (5), Eq. (3) is obtained.

Two approximations have been made in deriving Eq. (3) from Eq. (1). In the first approximation

$$\frac{1-j}{2}+j\left(\frac{b\lambda}{2}\right)^{\frac{1}{2}}\frac{1}{\pi x}\exp\left(\frac{-j\pi x^2}{b\lambda}\right)$$

was substituted for

$$C\left[\left(\frac{2}{b\lambda}\right)^{\frac{1}{2}}x\right]-jS\left[\left(\frac{2}{b\lambda}\right)^{\frac{1}{2}}x\right].$$

Table I compares the true and approximate expressions at each end of the ranges over which the approximation has been used.

The second approximation was for E_1 and is at least as good as the one used for Fresnel's integrals except in the neighborhood of $\alpha=\sin^{-1}(c/2l)$, where $L_1=0$.

Effect of Hydrostatic Pressure in an Electron Beam on the Operation of Traveling-Wave Devices*

PHILIP PARZEN AND LADISLAS GOLDSTEIN
Federal Telecommunication Laboratories, Inc., Nutley, New Jersey

(Received August 11, 1950)

Small velocity spreads in an electron beam appear to cause a decrease in gain and noise figure of a traveling-wave tube. The actual decrease depends on a quantity μ, which can be interpreted physically as a measure of the ratio of the coupling of the electrons to each other via the hydrostatic pressure to the coupling of the electrons to the external wave guide circuit. Thus, physically, the effect of the velocity spread is to introduce further interactions among the electrons, in which the external circuit does not take part. This interaction may be thought of as a hydrostatic pressure similar to that in a liquid or gas.

THE macroscopic properties of a gas can be calculated from the distribution function $f(x, u, t)$, where $fdxdu$ is the number of particles having their position between x and $x+dx$ and their velocity between u and $u+du$. This function can be determined from the Boltzmann transport equation which, if written in one dimension, is

$$(\partial f/\partial t)+u(\partial f/\partial x)+a\partial f/\partial u=\partial f/\partial t|_{\text{collision}}, \quad (1)$$

where a is the acceleration of a particle. Equation (1) is the mathematical statement of the conservation of the number of particles in detail along a flow line.

The mean (drift) velocity v and total particle density n are given by

$$n=\int f du, \quad (2)$$

$$v=(1/n)\int uf(u)du, \quad (3)$$

and only velocities in the x direction are considered.

As shown by Chapman and Cowling,[1] the integrals of Eq. (1), which express the conservation of matter on the whole and momentum, eliminate the collision term and also determine the variation of the total particle density and mean velocity.

These expressions are

$$(\partial n/\partial t)+\partial(nv)/\partial x=0, \quad (4)$$

$$dv/dt=a-\partial P/nm\partial x, \quad (5)$$

where m is the mass of a particle and P is the hydrostatic pressure.

The evaluation of P may be carried out in terms of an equivalent temperature T of the distribution function. Thus,

$$kT/2=(1/n)\int \tfrac{1}{2}m(u-v)^2 f du, \quad (6)$$

$$P=nkT, \quad (7)$$

k being Boltzmann's constant.

We see from Eq. (5) that the effect of the distribution on the mean velocity is to introduce an additional term in the force equation. The advantage of working with these integral expressions is that they constitute two simultaneous differential equations for n and v in those flow processes in which T is constant. This constitutes a considerable simplification in the solution of flow problems in a gas.

A, perhaps, trivial example of such a process is that of isothermal flow in a gas. The more important example of an approximately isothermal process that we wish to discuss here is that of interaction between an electron gas and an electromagnetic wave, in which the determination of n and v are of paramount importance.

It is well known that electromagnetic waves propagate in electron beams[2,3] and under certain conditions, which occur for example in the traveling-wave tube[4] and in the double-electron-beam tube,[5] there exists amplification of the electromagnetic wave. In these problems, the calculation of gain depends on the change in velocity, charge density, and current density by the perturbing electromagnetic traveling wave. The magnitude of this change will also depend on the distribution function f. We shall now calculate this for the one-dimensional case of an electron beam traveling in an electromagnetic field characterized by an electric field E in the same direction. Thus

$$E=\bar{E}\exp(j\omega t-\Gamma x), \quad (8)$$

$$a=-e\bar{E}\exp(j\omega t-\Gamma x)/m, \quad (9)$$

where the tilde indicate an alternating-current value.

The electron beam is considered as an electron gas with a direct-current distribution function $f_0(u)$ and drift velocity u_0. In accordance with the usual small-signal approximations, we assume that with the perturbing electric field

$$\left.\begin{aligned} v&=u_0+\tilde{v}\exp(j\omega t-\Gamma x)\\ n&=n_0+\tilde{n}\exp(j\omega t-\Gamma x)\\ T&=T_0+\tilde{T}\exp(j\omega t-\Gamma x)\\ J\ (\text{current density})&=J_0+\tilde{J}\exp(j\omega t-\Gamma x)\\ J_0&=-en_0u_0\\ \tilde{J}&=-e(n_0\tilde{v}+\tilde{n}u_0)\end{aligned}\right\}. \quad (10)$$

* This work was sponsored by the Signal Corps Engineering Laboratories.

[1] Chapman and Cowling, *The Mathematical Theory of Non-Uniform Gases* (Cambridge University Press, London, 1939), p. 51.

[2] W. C. Hahn, Gen. Elec. Rev. **42**, 258 (1939).

[3] S. Ramo, Proc. Inst. Radio Engrs. **27**, 757 (1939).

[4] J. R. Pierce, Proc. Inst. Radio Engrs. **35**, 114 (1947).

[5] A. V. Haeff, Proc. Inst. Radio Engrs. **37**, 4 (1949).

Substituting these into Eqs. (4), (5), and (7) and retaining only first-order terms, we have

$$j\omega\tilde{v}-\Gamma u_0\tilde{v}=-(e\tilde{E}/m)+\Gamma(kT_0\tilde{n}/mn_0)+\Gamma(k\tilde{T}/m), \quad (11)$$

$$j\omega\tilde{n}-\Gamma(n_0\tilde{v}+\tilde{n}u_0)=0. \quad (12)$$

We shall now assume that the process is isothermal. Thus, $\tilde{T}=0$. Under these conditions,

$$\tilde{J}=j\beta I_0\tilde{E}/2V_0[(-\Gamma+j\beta)^2-\alpha\Gamma^2], \quad (13)$$

$$\tilde{v}=-u_0(j\beta-\Gamma)\tilde{E}/2V_0[(j\beta-\Gamma)^2-\alpha\Gamma^2], \quad (14)$$

$$\tilde{n}=-n_0\Gamma\tilde{E}/2V_0[(j\beta-\Gamma)^2-\alpha\Gamma^2], \quad (15)$$

where $\beta=\omega/u_0$,

$$I_0=en_0u_0=\text{direct-current density}, \quad (16)$$

$V_0=$ voltage corresponding to u_0, and

$$\alpha=kT_0/mu_0^2,$$

which is a measure of the random energy to the directed energy. For $\alpha=0$, these equations reduce to those quoted elsewhere.[4,5]

To calculate the interaction between the electron beam and the electromagnetic field, it is necessary to add to these equations the circuit equations. These are obtained from Maxwell's equations and are determined by the geometry of the surrounding wave guide system. We shall now give these calculations for the traveling-wave tube, using the method of Pierce[4] to obtain the circuit equation for a filamentary electron beam.

EFFECT OF VELOCITY SPREAD IN A TRAVELING-WAVE TUBE WITH A FILAMENTARY ELECTRON BEAM

Here $n=$ number of electrons per unit length and $\tilde{J}=$ total current.

According to Pierce,[4] the relation between alternating current $\tilde{J}$ and the alternating electric field $\tilde{E}$ on the axis is

$$\tilde{E}=\tilde{J}\Gamma_0/\psi_0(\Gamma^2-\Gamma_0^2) \quad (17)$$

if one considers only one mode; Γ_0 and ψ_0 are the propagation constant and admittance, respectively, of the wave guide in the absence of the beam.

Substituting this into Eq. (13), we have the following equation for Γ:

$$\Gamma^2-\Gamma_0^2=j2\beta^3C^3\Gamma_0/(-\Gamma+j\beta)^2-\alpha\Gamma^2. \quad (18)$$

$C=$ Pierce coupling coefficient $=(I_0/\psi_0\beta^2 4V_0)^{\frac{1}{3}}$. (19)

We shall calculate Γ for the synchronous case only. Thus,

$$\Gamma_0=j\beta.$$

Let

$$\Gamma=j\beta-\delta; \quad \delta\ll\beta. \quad (20)$$

Thus, to first-order terms in δ, Eq. (19) becomes

$$\delta^3+\alpha\beta^2\delta=-j\beta^3C^3. \quad (21)$$

Now let $\delta=j\beta CX$ and $\mu=\alpha/C^2$, hence,

$$X^3-\mu X-1=0. \quad (22)$$

As is well known, an amplified wave corresponds to a complex root of Eq. (22).

Let $G(\mu)=$ gain of the growing wave in decibels per centimeter for any μ. Hence

$$G(\mu)/G(0)=2X_i/\sqrt{3}, \quad (23)$$

X_i being the imaginary part of a root of Eq. (22).

$G(0)$ corresponds to the case of a monochromatic beam. In Fig. 1, $G(\mu)/G(0)$ is plotted against μ. It is seen that the gain is reduced by the velocity spread in the beam. Let us consider the following example: $kT_0=0.10$ volt (corresponding to a cathode temperature $\approx 1200°$K),† $C=0.015$, and $V_0=500$ volts. Then $\mu=1$ and the reduction in gain is about 65 percent.

Thus, considerable reductions in gain from that given by the Pierce equation may be ascribed to a velocity spread in the electron beam.

EFFECT OF VELOCITY SPREAD ON NOISE FIGURE IN TRAVELING-WAVE TUBES

Here, it is important to know the magnitude of the field in each mode that is excited by an initial input of electric field E, velocity modulation $\tilde{v}$, or current modulation $\tilde{J}$. From Eqs. (13), (14), and (15), the initial conditions are

$$\Sigma E_i=E, \quad (24)$$

$$\Sigma(j\beta-\Gamma_i)E_i/(j\beta-\Gamma_i)^2-\alpha\Gamma_i^2=-2V_0\tilde{v}/u_0, \quad (25)$$

$$\Sigma[E_i/(j\beta-\Gamma_i)^2-\alpha\Gamma_i^2]=2V_0\tilde{J}/j\beta I_0, \quad (26)$$

for $\alpha=0$, these are the same conditions as those of Pierce.

If α is assumed to be small, these reduce to

$$\Sigma e_i[1-2j\alpha(\beta/\delta_i)+\alpha(\beta/\delta_i)^2]=E, \quad (24a)$$

$$\Sigma e_i/\delta_i=-2V_0\tilde{v}/u_0, \quad (25a)$$

$$\Sigma e_i/\delta_i^2=2V_0\tilde{J}/j\beta I_0, \quad (26a)$$

where

$$\delta_i=j\beta-\Gamma_i \quad \text{and} \quad e_i=E_i/[1-2j\alpha(\beta/\delta_i)+\alpha(\beta/\delta_i)^2].$$

The gaining wave E_1 with a corresponding root X_1 of Eq. (22) is

$$E_1=[1-2j\alpha(\beta/\delta_i)+\alpha(\beta/\delta_i)^2]\times\frac{E+\tilde{v}(2V_0/u_0)(\delta_2+\delta_3-2j\alpha\beta)+(2V_0/j\beta I_0)(\delta_2\delta_3-\alpha\beta^2)\tilde{J}}{(1-\delta_2/\delta_1)(1-\delta_3/\delta_1)}. \quad (27)$$

We shall now compute the noise figure F for the following two cases using the method of Pierce.

† It can be shown with the aid of Eq. (40) that the direct-current process for rectilinear electron beams of small current density is essentially isothermal.

Case 1

Beam enters the wave guide with only current-modulation fluctuations.

Here,

$$F(\mu)/F(0)=|\delta_2\delta_3-\alpha\beta^2|^2/\beta^4C^4 = |1+X_1\mu|^2/|X_1|^2, \quad (28)$$

$$F(0)=2\Gamma^2eV_0C/kT. \quad (29)$$

In Fig. 1, we have plotted $F(\mu)/F(0)$ against μ. It is seen that the noise figure is reduced by the velocity spread in the beam.

Case 2

Beam enters the wave guide with only velocity modulation $\tilde{v}$.

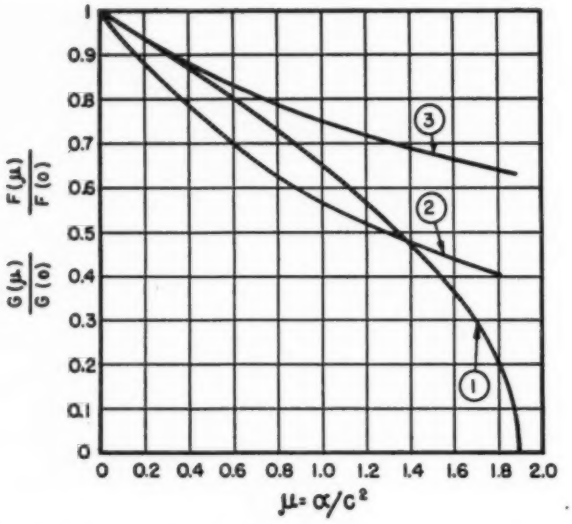


FIG. 1. Reduction in gain and noise figure in a traveling-wave tube as a result of velocity spread. Plotted against μ, are relative values of (1) gain in decibels per centimeter, (2) noise figure for current-modulation fluctuations, and (3) noise figure for velocity-modulation fluctuations.

Here,[6]

$$\langle v^2\rangle_{kT}=4\eta kT_c(1-\tfrac{1}{4}\pi)\Delta f/I_0, \quad (30)$$

where $\eta=e/m$ and T_c=cathode temperature in degrees K.

$$F(\mu)/F(0)=|\delta_2+\delta_3-2j\alpha\beta|/\beta^2C^2=|X_1|^2, \quad (31)$$

$$F(0)=2(1-\tfrac{1}{4}\pi)T_c/TC, \quad (32)$$

where T=ambient temperature in degrees K.

In Fig. 1, $F(\mu)/F(0)$ is plotted against μ. Here, too, the noise figure is reduced by velocity spread.

PARTIAL JUSTIFICATION OF $\tilde{T}=0$

We shall now determine under what conditions the alternating-current temperature $\tilde{T}$ may be neglected. We shall estimate this by deriving Eq. (13) by another method that deals with the distribution function f itself without regard to the concepts of pressure and temperature, which may be derived from f.

Let $f_0(u)du$ be the direct-current distribution function. Then, from the Boltzmann transport equation, neglecting collisions and making the usual small-signal approximations, the alternating-current density $d\tilde{J}$ due to the electrons whose speed lies between u and $u+du$ is given by

$$d\tilde{J}=j\omega\eta eEf_0(u)du/(j\omega-\Gamma u)^2. \quad (33)$$

This can also be obtained by a derivation similar to that of Pierce.[7] Hence,

$$\tilde{J}=j\omega\eta eE\int f_0(u)du/(j\omega-\Gamma u)^2=j\omega\eta eEP. \quad (34)$$

Now let $\Delta=u-u_0$ and assume that f_0 is such that $f_0(\Delta)\neq 0$ only for $\Delta<\Delta_0$.

Thus, we are dealing only with very small velocity spreads. Hence,

$$P=\int_{-\Delta_0}^{\Delta_0} f_0(\Delta)d\Delta/(j\omega-\Gamma u_0-\Gamma\Delta)^2. \quad (35)$$

Let us further assume that

$$|j\omega-\Gamma u_0|\gg|\Gamma\Delta_0|. \quad (36)$$

Then

$$P=\frac{1}{(j\omega-\Gamma\mu_0)^2} \times\int f_0(\Delta)\left[1-\frac{2\Gamma\Delta}{j\omega-\Gamma\mu_0}+\frac{3\Gamma^2\Delta^2}{(j\omega-\Gamma\mu_0)^2}\right]d\Delta, \quad (37)$$

by expanding the integrand of Eq. (35) up to terms in Δ^2.

Now

$$\left.\begin{aligned}&\int f_0(\Delta)d\Delta=n_0\\ &\int f_0(\Delta)\Delta d\Delta=0\\ &\int f_0(\Delta)\Delta^2d\Delta=n_0kT_0/m \text{ by Eq. (6)}\end{aligned}\right\}. \quad (38)$$

Hence,

$$\tilde{J}=[j\beta I_0\tilde{E}/2V_0(-\Gamma+j\beta)^2][1+3\alpha\Gamma^2/(-\Gamma+j\beta)^2], \quad (39)$$

which is equivalent to Eq. (13) for small α/C^2 except for a factor of 3. The requirement stated in Eq. (36) is also valid for small α/C^2.

This derivation, however, is not correct, in that it neglects the effect of collisions. There are two types of

[6] A. J. Rack, Bell System Tech. J. **17**, 592 (1938).

[7] J. R. Pierce, J. Appl. Phys. **19**, 231 (1948).

collisions to be considered. The collisions of electrons with gas atoms or positive ions may be neglected if the frequency of the electromagnetic wave is much greater than the collisional frequency. This will always be true for sufficiently high frequencies and sufficiently low gas pressures. The collisions between electrons because of their coulombian law of interaction may not be neglected. It is the electron-electron interaction that will determine whether the process is isothermal.

While one may not hope to solve the Boltzmann equation with this collision term, it is possible to use a method of successive approximations[8] that depends on the fact that f is completely determined by a knowledge of certain macroscopic quantities that are functions of position and time only. These are n, v, and T. The variation of n and v is given by Eqs. (4) and (5). The equation[1] governing the variation of T, which is equivalent to the conservation of energy, is

$$dT/dt=-2T(\partial v/\partial x)-2\partial q/kn\partial x. \tag{40}$$

The random energy flow per unit area per unit time is

$$q=nm\int \tfrac{1}{2}(u-v)^3 f du. \tag{41}$$

In this method of successive approximations, the first approximation is one for which $q=0$ and f is essentially maxwellian. For the second approximation, which is the first to depend on the nature of the interaction or collisions among the particles,

$$q=-\lambda\partial T/\partial x, \tag{42}$$

where λ, the coefficient of thermal conductivity, is a function of the collision forces, n, and T. Thus, to this order of approximation, n, v, T may be determined from Eqs. (4), (5), (40), (41), and (42). Repeating the steps of the small-signal approximation as previously gives the alternating-current density $\bar{J}$ as

$$\bar{J}=j\beta I_0 E/2V_0[(j\beta-\Gamma)^2-\alpha\Gamma^2(1+2/r)], \tag{43}$$

with

$$r=1+2\lambda\Gamma^2/kn_0(j\omega-\Gamma\mu_0), \tag{44}$$

where, in accordance with the small-signal approximations, λ is now a constant and is computed at n_0, T_0. There is a similar change in the denominator in the expressions for $\bar{n}$ and $\bar{v}$.

Now, for an adiabatic process, $\lambda=0$ and $r=1$; and for an isothermal process, $\lambda=\infty$ and $r=\infty$.

Thus, as λ goes from 0 to ∞, this is the equivalent of the process going from adiabatic to isothermal. Also, it is seen from Eq. (43) that the effect of an adiabatic process over that of an isothermal process is to triple the value of α. Thus, it is possible to say that, for the actual physical process, the effective α will always lie somewhere between α and 3α. It is seen in general that for $r\gg 2$, the process is essentially isothermal. Now λ can be computed for the case of electron-electron interaction from an equation given by Chapman and Cowling.[9] For $n_0=10^9$ and $T_0=1000°$K, $\lambda=18$ cgs units. Taking the values of Γ computed previously on the assumption of an isothermal process and substituting them into r, r is given approximately by

$$r=1+2\lambda\beta^2/kn_0\omega CX. \tag{45}$$

For practical values, $\beta>10$ and $\omega=2\pi\times 5\times 10^9$, $X\approx 1$; $r\approx 100$. Hence, in the traveling-wave tube, the process is isothermal. In general, whether the process is isothermal can be ascertained only by computing r for each process and geometry under consideration. Furthermore, it is seen that the analysis used in deriving Eq. (39) tacitly assumes an adiabatic process in that it neglects the coulombian collisions among particles.

CONCLUSIONS

Small velocity spreads in an electron beam appear to cause a decrease in gain and noise figure of a traveling-wave tube. The actual decrease depends upon a quantity μ, which can be interpreted physically as a measure of the ratio of the coupling of the electrons to each other via the hydrostatic pressure to the coupling of the electrons to the external wave guide circuit. Thus, physically, the effect of the velocity spread is to introduce further interactions between the electrons, in which the external circuit does not take part. This interaction may be thought of as a hydrostatic pressure,[10] similar to that in a liquid or gas.

[8] See reference 1, p. 107.

[9] See reference 1, p. 179.

[10] W. C. Hahn, Proc. Inst. Radio Engrs. **36**, 1115 (1948).

On Stability of Linear Varying-Parameter Systems

LOTFI A. ZADEH
Columbia University, New York, New York
(Received August 25, 1950)

A linear varying-parameter system is defined to be *stable* if and only if every bounded input produces a bounded output. It is shown that a necessary and sufficient condition for stability is that the impulsive response of the system $W(t, \tau)$ should be integrable (considered as a function τ) for all t. From this result and the fact that the system function $H(s; t)$ is the laplace transform of $W(t, \tau)$, it is deduced that a necessary condition for stability of a linear varying-parameter system is that the system function $H(s; t)$ should be analytic and bounded in the right half and on the imaginary axis of the s-plane for all t. This result represents a generalization of the familiar frequency domain criterion which is commonly used in connection with fixed systems. The generalized criterion is applied to the investigation of stability of a variable feedback system.

INTRODUCTION

A LINEAR varying-parameter system is essentially a physical system in which one or more parameter values are specified[1] functions of time. In the case of "lumped-constant" systems of this type, the relations describing the behavior of the system consist of linear differential equations with time-varying coefficients. More generally, however, these relations may involve time-dependent Heaviside operators[2,3] of the form $H(p; t)$, where $H(p; t)$, in general, is not a polynomial in p. The pertinent properties of these operators will be discussed later in the paper.

The classical theory[4] of stability of variable systems is based on the study of boundedness and asymptotic behavior of solutions of linear differential equations. The theory outlined in the present paper, however, is based on a different point of view. Specifically, our approach consists essentially in investigating the stability of a system through an examination of the behavior of its system function in the complex frequency domain. This approach is, in effect, a generalization of the conventional frequency domain technique which is widely used in connection with fixed systems.

Preliminary to the discussion of stability of variable systems, it is useful to summarize the basic definitions and relations which will be needed later in this paper. These are given in the following section.

VARIABLE SYSTEMS

Basic Notation and Assumptions

It will always be assumed, unless specified to the contrary, that the system is unexcited, though not necessarily fixed, prior to the application of the input. The input is denoted by $u(t)$ and the response to $u(t)$ (i.e., the output of the system) is denoted by $v(t)$. The system itself is denoted by N. The system need not be of the "lumped-constant" type.

Impulsive Response

The *impulsive response* of N is a function $W(t, \tau)$ which represents the response of N at t to a unit impulse (Dirac's δ-function) applied at $t-\tau$. The response of N to an arbitrary input $u(t)$ may be expressed in terms of $W(t, \tau)$ as

$$v(t)=\int_0^\infty W(t, \tau)u(t-\tau)d\tau. \tag{1}$$

In mathematical terminology, $W(t, \tau)$ represents an *integral operator* relating $v(t)$ and $u(t)$.

System Function

The system function of N is defined as a function $H(j\omega; t)$ such that $H(j\omega; t)e^{j\omega t}$ represents the response of N to $e^{j\omega t}$. In other words,

$$H(j\omega; t)=v(t)/u(t)\,|_{u(t)=\exp(j\omega t)}. \tag{2}$$

The system function may also be written as $H(s; t)$, s standing for the complex frequency $s=\sigma+j\omega$. The response of N to an arbitrary input $u(t)$ may be expressed in terms of $H(j\omega; t)$ in several different forms of which the most useful for the purposes of this paper is the operational form

$$v(t)=H(p; t)u(t). \tag{3}$$

In using this relation, the time-dependent operator $H(p; t)$ should be treated as an ordinary[5] Heaviside operator, that is, the variable t in $H(p; t)$ should be treated as if it were a constant parameter.

Relation between the System Function and the Impulsive Response

The system function $H(s; t)$ and the impulsive response $W(t, \tau)$ constitute a laplace transform pair in-

[1] More generally, the parameter values may be random functions of time. Systems of this type are considered in Proc. Inst. Radio Engrs. **38**, 1342 (1950).
[2] L. A. Zadeh, J. Appl. Phys. **21**, 1171 (1950).
[3] L. A. Zadeh (to be published in J. Math. Phys).
[4] R. Bellman, "*A Survey of the Theory of the Boundedness, Stability, and Asymptotic Behavior of Solutions of Linear and Nonlinear Differential and Difference Equations*" (ONR, Washington, D. C., 1949).
[5] It should be noted that whenever the input $u(t)$ does not vanish for negative t, $H(p; t)$ should be treated as a *bilateral* operator. This means that in evaluating $H(p; t)u(t)$ one should use the Fourier transform or the bilateral Laplace transform of $u(t)$ and not the unilateral Laplace transform of $u(t)$.

volving t as a parameter. Thus,

$$H(s;t)=\mathfrak{L}\{W(t,\tau)\} \tag{4}$$

and

$$W(t,\tau)=\mathfrak{L}^{-1}\{H(s;t)\}, \tag{5}$$

where $\mathfrak{L}$ represents the operation of laplace transformation with respect to τ.

Product Relation

Let an operator $H_1(p;t)$ operate on $u(t)$ and let $H_2(p;t)$ operate on $H_1(p;t)u(t)$. The over-all result is equivalent to operating on $u(t)$ with an operator $H_3(p;t)$ which is called the *product* of $H_2(p;t)$ and $H_1(p;t)$. This operator is given by the operational relation

$$H_3(s;t)=H_2(p+s;t)H_1(s;t), \tag{6}$$

where $H_1(s;t)$ plays the role of a function of time involving s as a parameter. Equation (6) may be written more conveniently in a symbolic form

$$H_3=H_2*H_1, \tag{7}$$

where the operation $*$ is distributive, associative, and, in general, non-commutative. The symbol $*$ may be treated in much the same manner as the usual algebraic product and reduces to the latter whenever the operators in question do not involve either p or t.

As should be expected, there exists a close connection between the product of system functions as defined above and the convolution of impulsive responses. Thus, let $W_1(t,\tau)$, $W_2(t,\tau)$, and $W_3(t,\tau)$ be the impulsive responses associated, respectively, with $H_1(p;t)$, $H_2(p;t)$, and $H_3(p;t)$. From Eq. (1) and the definition of the product, it follows that if H_3 is the product of H_2 and H_1, then W_3 is related to W_2 and W_1 by the convolution integral

$$W_3(t,\tau)=\int_0^\infty W_2(t,\tau')W_1(t-\tau',\tau-\tau')d\tau'. \tag{8}$$

For fixed systems Eqs. (6) and (8) reduce to the familiar relations

$$H_3(s)=H_2(s)H_1(s), \tag{9}$$

and

$$W_3(\tau)=\int_0^\infty W_2(\tau')W_1(\tau-\tau')d\tau'. \tag{10}$$

The Inverse of a Time-Dependent Operator

The inverse of $H(p;t)$, or simply H, is defined as an operator H^{-1} such that the product (in the sense of Eq. (6)) of H and H^{-1} is equal to unity. Thus,

$$H*H^{-1}=H^{-1}*H=1. \tag{11}$$

The properties of system functions, as outlined here, will now be applied to the investigation of stability of linear varying-parameter systems.

STABILITY

As the definition of stability in the case of variable systems, we shall use an extended form of a definition formulated by Hurewicz, James, and Weiss[6] for fixed systems. In its extended form the definition reads:

(*A*) *A variable system N is said to be stable if and only if every bounded input produces a bounded output. Thus, if some bounded input produces an unbounded output, the system is unstable.*

Our objective is to establish a connection between the stability of a variable system and the behavior of its system function $H(s;t)$ in the complex frequency plane (s-plane). As a preliminary, we shall prove the following statement:

(*B*) *A variable system N is stable if and only if its impulsive response $W(t,\tau)$ is integrable with respect to τ for all values of t, i.e.,*

$$\int_0^\infty |W(t,\tau)d\tau<\infty \quad \text{for all } t. \tag{12}$$

In mathematical terminology, this means that $W(t,\tau)$ (considered as a function of τ) should belong to L over $(0,\infty)$ for all t.

To prove the sufficiency of (*B*), let it be assumed that $W(t,\tau)$ satisfies Eq. (12) and that the input is bounded, i.e., $|u(t)|\leqslant M<\infty$. Then, Eq. (1) gives

$$|v(t)|\leqslant\int_0^\infty |W(t,\tau)|\,|u(t-\tau)|d\tau \tag{13}$$

or

$$|v(t)|\leqslant M\int_0^\infty |W(t,\tau)|d\tau, \tag{14}$$

and, hence, in view of Eq. (12),

$$|v(t)|<\infty, \tag{15}$$

which shows that if $W(t,\tau)$ is integrable, then the response of N to any bounded input is bounded.

To prove the necessity of (*B*), it is sufficient to show that if $W(t,\tau)$ is not integrable for some t, say $t=t_0$, then there is some bounded input which produces an unbounded output. Such an input can be constructed as follows.

For $t=t_0$, let $u(t_0-\tau)$ be a bounded function which is equal to 1 wherever $W(t_0,\tau)$ is positive, and is equal to -1 wherever $W(t_0,\tau)$ is negative. From Eq. (1), the response to this input at $t=t_0$ is

$$v(t_0)=\int_0^\infty |W(t_0,\tau)|d\tau, \tag{16}$$

and if $W(t_0,\tau)$ is not integrable, then $v(t_0)$ is unbounded. Thus, it follows that if Eq. (12) is not satisfied, then

[6] James, Nichols, and Phillips, *Theory of Servomechanisms* (McGraw-Hill Book Company, Inc., New York, 1947), chapters 2 and 5. The proofs of (*B*) and (*C*) are essentially extensions to variable systems of some of the proofs given in these two chapters.

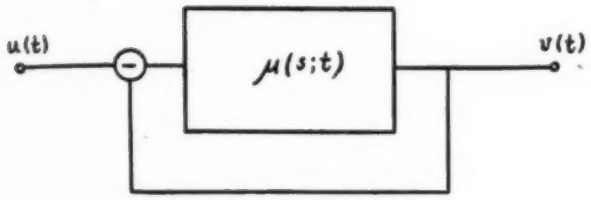


FIG. 1. A variable feedback system.

there is some bounded input which produces an unbounded output, and, hence, the system is unstable. This establishes the necessity of (B).

Turning now to $H(s;t)$, we note from Eq. (4) that $H(s;t)$ is the Laplace transform of $W(t,\tau)$ with respect to τ. As is well known,[7] the Laplace transform of an integrable function is analytic in the right half plane and on the imaginary axis. Hence, from (B) it follows that:

(C) *The system function of a stable variable system is analytic and bounded in the right half and on the imaginary axis of the s-plane for all t.*

In order to simplify the terminology, the system function of a stable variable system will be referred to simply as a *stable* system function. From (B) it follows that a stable system function is one whose inverse laplace transform is integrable (as a function of τ) for all values of t.

The analyticity and boundedness of $H(s;t)$ in the right half plane and on the imaginary axis is a necessary but not a sufficient condition for stability of the system function. Thus, $H(s;t)$ may satisfy (C) and yet $W(t,\tau)$ might not be integrable. To ensure the integrability of $W(t,\tau)$ and, hence, the stability of the system, certain additional conditions must be imposed on the behavior of $H(s;t)$ along the imaginary axis. It can be shown that a sufficient but not a necessary condition of this kind is that $|H(s;t)|^2$ and $|\partial H(s;t)/\partial s|^2$ should be integrable along the imaginary axis for all t. However, in practice it is generally unnecessary to be concerned with these supplementary conditions. Thus, for all practical purposes we may state that:

(D) *A necessary and (practically) sufficient condition for stability of a variable system N is that its system function $H(s;t)$ should be analytic and bounded in the right half and on the imaginary axis of the s-plane for all t.*[8]

This criterion represents a generalization of the familiar stability criterion involving the analyticity of $H(s)$ (the system function of a fixed system) in the right half of the complex frequency plane.

In applying the above criterion to the investigation of stability of variable systems, the following property of system functions is frequently found useful.

Consider two variable systems N_1 and N_2 which are connected in tandem, and let N_3 represent the composite system. From (A) it follows at once that if N_1 and N_2 are stable systems, then so is N_3. Now let H_1 and H_2 denote the system functions of N_1 and N_2. The system function of N_3 is H_1*H_2 or H_2*H_1 (depending on whether N_2 precedes N_1 or vice versa). Thus, we are led to the following conclusion:

(E) *If H_1 and H_2 are stable system functions, then so are the products H_1*H_2 and H_2*H_1.*

As a simple application of the above results, consider a feedback system (Fig. 1) whose forward gain is

$$\mu(s;t)=(a+b\cos t)/(1+Ts), \tag{17}$$

where a, b, and T are specified constants ($T>0$). The feedback factor β is assumed equal to unity. The problem is to determine whether the system is stable or unstable.

The input $u(t)$, the output $v(t)$, and the gain $G(s;t)$ of the system under consideration are related to each other by the operational relation

$$v(t)=G(p;t)u(t). \tag{18}$$

Thus, our problem is essentially that of determining whether or not $G(s;t)$ is a stable system function.

The gain $G(s;t)$ (or simply G) can readily be expressed in terms of the forward gain $\mu(s;t)$ (or simply μ). We have

$$G=(1+\mu)^{-1}*\mu \tag{19}$$

which is a generalization of the familiar relation

$$G=\mu/(1+\mu). \tag{20}$$

Referring to Eq. (17), it is seen that the assumed expression for $\mu(s;t)$ represents a stable system function. Hence, from (E) it follows that the stability or instability of G depends on that of the factor $(1+\mu)^{-1}$. Calling this factor F and making use of Eq. (17), we find

$$\begin{aligned} F&=\left(\frac{Tp+1+a+b\cos t}{1+Tp}\right)^{-1}\\ &=\left[(Tp+1+a+b\cos t)*\frac{1}{(1+Tp)}\right]^{-1}\\ &=(1+Tp)*(Tp+1+a+b\cos t)^{-1}\\ &=(1+Tp)*J^{-1}, \end{aligned} \tag{21}$$

where

$$J=Tp+1+a+b\cos t. \tag{22}$$

Thus, in the final analysis the stability of $G(s;t)$ depends on that of J^{-1}.

To obtain the inverse of J, it is necessary to solve the differential equation

$$J*J^{-1}=1, \tag{23}$$

which in explicit form reads

$$[T(p+s)+1+a+b\cos t]J^{-1}(s;t)=1. \tag{24}$$

[7] See, for instance, G. Doetsch, *Theorie und Anwendung der Laplace Transformation* (Verlag. Julius Springer, Berlin, 1937).

[8] It is evident that whenever one is interested only in the behavior of the system for some $t>t_1$, then in statements (B), (C), and (D), the words "for all t" should be replaced by "for $t>t_1$."

This equation is of the first order and its solution yields

$$J^{-1}(s;t)=\frac{1}{T}e^{-b\cos t/T}$$

$$\times\int_0^\infty \exp\left\{-\frac{(1+a+sT)}{T}\tau+\frac{b}{T}\cos(t-\tau)\right\}d\tau. \quad (25)$$

Inspection of this expression shows that J^{-1} is stable or unstable according as $(1+a)$ is positive or negative. In view of Eqs. (19) and (21), $G(s;t)$ is stable or unstable under the same conditions. Consequently, we may conclude that:

The feedback system under consideration is stable for $(1+a)>0$ and is unstable for $(1+a)\leqslant 0$.

CONCLUDING REMARKS

The frequency domain approach outlined in this paper is essentially an extension to variable systems of a frequency domain technique which is extensively used in investigations of stability of fixed systems. In such systems this technique has proved extremely useful. It might also prove useful in the case of variable systems.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 4 APRIL, 1951

The Effect of a Grounded Slab on the Radiation from a Line Source*

C. T. TAI
Stanford Research Institute, Stanford, California
(Received August 14, 1950)

A theoretical investigation, based upon the method of fourier transforms, is made of the effect of a grounded dielectric slab on the radiation from a line source. It is shown that the principal part of the electric field above the dielectric slab can be found either by evaluating the resultant contour integral, using the saddle-point method of integration, or by evaluating the integral containing the tangential H-field along the interface. The radiation pattern of the principal field is found to be identical with the resultant field of a direct ray and a reflected ray that can be obtained by a simple consideration of geometrical optics. The condition of the existence of propagating modes in the dielectric slab has been discussed in detail. When the slab is thick enough to support a propagating mode, in addition to the space wave, a surface wave will appear in the neighborhood of the interface. The latter attenuates rapidly as the point of observation is moved away from the interface. The deformation of the path of integration corresponding to different angles of observation has been displayed graphically both in the original h-plane and the transformed τ-plane.

I. INTRODUCTION

THE problem which will be considered in this report is illustrated graphically in Fig. 1, in which I is a line source of current located at $x=b$ above the ground plane, which is assumed to be perfectly conducting and of infinite extent in the $y-z$ plane. A dielectric slab of thickness a and of index of refraction equal to n lies on the ground plane. It is well known that without the dielectric slab, the radiation pattern of the E-field is proportional to $\sin(2\pi b/\lambda\cos\theta)$. The purpose of this investigation is to find out the change in the radiation pattern which results from the presence of the dielectric slab. Analytically, the present problem can be considered as a two-dimensional version of a similar problem formulated by Van der Pol and Bremmer.[1] It is also closely related to the problem of the parallel plate dielectric wave guide discussed by Whitmer.[2] The contour integral which represents the resultant field, obtained by Van der Pol and Bremmer, was discussed only qualitatively. Whitmer evaluated his contour integral by deforming the original path of integration so that there results a branch-cut integration and a number of residue integrations. The branch-cut integration was performed under the condition that the point of observation is a large distance from the source and close to the guide surface. The result does not provide enough information about the field distribution as a whole around the guide. In the present investigation, the problem will first be formulated using two different, but equivalent, methods. The resultant contour integral obtained by both methods will, in general, be evaluated by means of the saddle-point method of integration. For some special cases, a branch-cut integration must also be performed. It is shown that the principal part of the resultant field, which is derived by the saddle-point method of integration, can be obtained from the fourier transform of the tangential H-field, if the point of observation is not too near the interface. The limitation of the second method is discussed in detail. The resultant contour integral is finally evaluated for several distinct cases corresponding to different angles of observation.

* The research reported in this paper was made possible by the sponsorship of the Army Air Forces under Contract AF 19(122)78.

[1] B. Van der Pol and H. Bremmer, Phil. Mag. **24**, 825 (1937). See also J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill Book Company, Inc., New York, 1941), p. 590, ex. 12, or S. A. Schelkunoff, *Electromagnetic Waves* (D. Van Nostrand Company, Inc., New York, 1943), p. 428.

[2] R. M. Whitmer, Proc. Inst. Radio Engrs. **36**, 1105 (1948).

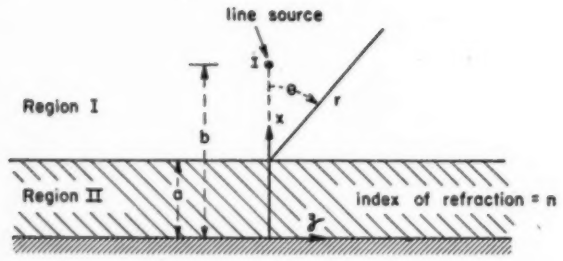


FIG. 1. A dielectric slab on a perfectly conducting plane with a line source of current above it.

II. FORMULATION

As mentioned in the introduction, the formulation of this problem will be carried out by two different, but equivalent, methods. The first method is the same as that used by Whitmer.[2] Its object is to find a composite Green's function which satisfies the boundary conditions that both the function and its normal derivative are continuous at the interface. The second method is to find two separate Green's functions for Regions I and II as shown in Fig. 1. These two Green's functions both satisfy the boundary condition that their normal derivatives vanish at the interface. The resultant E-field in each region then can be expressed as a line integral of the tangential H-field along the interface. The integral equation of the tangential H-field, obtained by matching the E-field at the interface, can be solved by an inverse fourier transform. The details are given below.

A. Method I

Consider the dielectric slab of depth a, and with index of refraction n which is placed on a perfectly conducting plane as shown in Fig. 1. The filament of current is at $x=b$, $z=0$. The electromagnetic field in this problem has three components, E_y, H_x, and H_z. Assuming that the magnitude of the current element, I, is numerically equal to $1/i\omega\mu$, the E-fields in Regions I and II satisfy, respectively, the following two equations,

$$(\nabla^2+k^2)E_1(x,z)=-\delta(x-b)\delta(z), \tag{1}$$

$$(\nabla^2+n^2k^2)E_2(x,z)=0, \tag{2}$$

where $k=2\pi/\lambda_{\text{air}}$ and δ denotes the Dirac delta-function. It is assumed that the time dependence is of the form, $e^{-i\omega t}$. This factor has been omitted in Eqs. (1) and (2). Let the fourier transforms of $E_1(x,z)$ and $E_2(x,z)$ be denoted by $\phi_1(x,h)$ and $\phi_2(x,h)$, or

$$\phi_1(x,h)=\int_{-\infty}^{\infty}E_1(x,z)e^{-ihz}dz, \tag{3}$$

$$\phi_2(x,h)=\int_{-\infty}^{\infty}E_2(x,z)e^{-ihz}dz, \tag{4}$$

then the fourier integrals

$$E_1(x,z)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\phi_1(x,h)e^{ihz}dh, \tag{5}$$

$$E_2(x,z)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\phi_2(x,h)e^{ihz}dh, \tag{6}$$

are solutions of Eqs. (1) and (2) provided that $\phi_1(x,h)$ and $\phi_2(x,h)$ satisfy the following equations:

$$\left(\frac{d^2}{dx^2}-\lambda^2\right)\phi_1(x,h)=-\delta(x-b), \tag{7}$$

$$\left(\frac{d^2}{dx^2}-\mu^2\right)\phi_2(x,h)=0, \tag{8}$$

where[3]

$$\left.\begin{aligned}\lambda^2&=h^2-k^2\\ \mu^2&=h^2-n^2k^2\end{aligned}\right\}. \tag{9}$$

In order that $\phi_1(x,h)$ be an outgoing wave for $x>b$, and $\phi_2(x,h)$ be vanishing at $x=0$, the solutions of Eqs. (7) and (8) must be of the following form:

$$\phi_1(x,h)=\begin{cases}A_1e^{-\lambda x}, & x\geqslant b, \quad (10)\\ A_2e^{\lambda x}+A_3e^{-\lambda x}, & b\geqslant x\geqslant a, \quad (11)\end{cases}$$

$$\phi_2(x,h)=A_4\sinh\mu x, \qquad a\geqslant x\geqslant 0, \tag{12}$$

where λ is assumed to be positive when $h>k$, and equal to $-i(k^2-h^2)^{\frac{1}{2}}$ when $h<k$. By imposing the boundary conditions similar to that discussed by Whitmer, the four arbitrary constants contained in Eqs. (10–12) can be determined. The final expression for $E_1(x,z)$ in the range $x>b$ is then given by the following integral

$$E_1(x,z)=\int_{-\infty}^{\infty}\frac{1}{2\pi\lambda}\times\left[\frac{\mu\sinh\lambda(b-a)+\lambda\cosh\lambda(b-a)\tanh\mu a}{\mu+\lambda\tanh\mu a}\right]\cdot e^{-\lambda(x-a)+ihz}dh. \tag{13}$$

The path of integration and the condition for the existence of the poles will be discussed in a later section. By a proper rearrangement of the terms, Eq. (13) can also be written in the form,

$$E_1(x,z)=\int_{-\infty}^{\infty}\frac{1}{4\pi\lambda}\times\left[e^{-\lambda(x-b)}-\left(\frac{\mu-\lambda\tanh\mu a}{\mu+\lambda\tanh\mu a}\right)e^{-\lambda(x+b-2a)}\right]e^{ihz}dh. \tag{14}$$

[3] The parameter λ must not be confused with the same notation used for the free space wavelength.

In order to give a physical interpretation of the two terms contained in Eq. (14), the integral representation of the free space Green's function of a two-dimensional wave equation must be introduced.

It is known from the theory of Green's functions[4] that the free space Green's function which satisfies a two-dimensional wave equation is given by

$$G(x-x', z-z')=(i/4)H_0^{(1)}(kr), \tag{15}$$

where $r=[(x-x')^2+(z-z')^2]^{\frac{1}{2}}$. The prime coordinates denote the position of the unit source, and the unprimed coordinates the point of observation. Using the integral representation of the Hankel function, one can write

$$G(x-x', z-z')=\begin{cases}\displaystyle\int_{-\infty}^{\infty}\frac{e^{-\lambda(x-x')+ih(z-z')}}{4\pi\lambda}dh, & x>x', \quad (16)\\ \displaystyle\int_{-\infty}^{\infty}\frac{e^{\lambda(x-x')+ih(z-z')}}{4\pi\lambda}dh, & x<x'. \quad (17)\end{cases}$$

With the help of Eqs. (16–17), the interpretation of Eq. (14) is relatively simple. The first term corresponds to the direct radiation from a line source at $x'=b$, and $z'=0$. The second term is due to the reflection from the dielectric slab. Consider the integral as a superposition of elementary plane waves, then the reflection coefficient for a typical elementary wave is given by

$$R=-\frac{\mu-\lambda\tanh\mu a}{\mu+\lambda\tanh\mu a}. \tag{18}$$

This concludes the derivation of the essential formulas by Method I.

B. Method II

The second method may seem to be more complicated; but as will be shown later, the principal part of the radiation field can be obtained in this formulation without a lengthy discussion of the contour integration.

Consider the same two differential equations given in Eqs. (1–2) which are defined in two separate Regions I and II with the common boundary at $x=a$. It is known[5] that the general solution of each differential equation can be expressed in terms of the boundary values of E and its normal derivative at an arbitrarily chosen closed surface with an appropriately chosen Green's function satisfying the same differential equation. Let the coordinate of the point of observation be denoted by r, and that of the source by r'. Then the field $E_1(r)$ for Region I is given by

$$E_1(r)=G_1(r, r')_{r'=r_0}+\int\left[G_1(r, r')\frac{\partial E_1(r')}{\partial n_1}-E_1(r')\frac{\partial G(r, r')}{\partial n_1}\right]dz', \tag{19}$$

where the following abbreviated notation for the Green's function is used,

$$G_1(r, r')=G_1(x, x', z-z'), \tag{20}$$

and r_0 denotes the coordinate of the line source, that is, $x'=b$, and $z'=0$. The closed surface was so chosen that the integral has to be evaluated only along the boundary $x=a$, since integration along the infinite semicircle in the upper plane contributes nothing to the closed line integral as a result of the radiation condition imposed upon the solution. Similarly, in Region II,

$$E_2(r)=\int\left[G_2(r, r')\frac{\partial E_2(r')}{\partial n_2}-E_2(r')\frac{\partial G_2(r, r')}{\partial n_2}\right]dz'. \tag{21}$$

The last integral has to be evaluated at both $x=0$, and $x=a$. If the following boundary conditions are imposed upon G_1 and G_2,

$$\partial G_1(r, r')/\partial n_1=-\partial G_1(x, x', z-z')/\partial x=0, \quad x'=a, \tag{22}$$

$$\partial G_2(r, r')/\partial n_2=\partial G_2(x, x', z-z')/\partial x=0, \quad x'=a, \tag{23}$$

and

$$G_2(r, r')=G_2(x, x', z-z')=0, \quad x'=0, \tag{24}$$

Eqs. (19) and (21) reduce to the following form

$$E_1(x, z)=G_1(x, b, z-0)-i\omega\mu\int_{-\infty}^{\infty}H_z(a, z')G_1(x, a, z-z')dz', \tag{25}$$

$$E_2(x, z)=i\omega\mu\int_{-\infty}^{\infty}H_z(a, z')G_2(x, a, z-z')dz'. \tag{26}$$

The appearance of the magnetic field H_z in Eqs. (25–26) is due to the identity

$$\partial E_y/\partial x=i\omega\mu H_z. \tag{27}$$

Since $H_z(x', y')$ is continuous across the boundary, $x'=a$, both E_1 and E_2 contain the same unknown quantity $H_z(a, z')$. To determine $H_z(a, z')$, the final boundary condition that E_y should be continuous at $x=a$ is imposed. If x is set equal to a in Eqs. (25–26), and the right-hand terms are equated, the following integral equation for $H_z(a, z)$ results:

$$i\omega\mu\int_{-\infty}^{\infty}H_z(a, z')[G_1(a, a, z-z')+G_2(a, a, z-z')]dz'=G_1(a, b, z-0). \tag{28}$$

Equation (28) can be solved for $H_z(a, z')$ by using the method of fourier transform. By multiplying Eq. (28) by e^{-ihz} and integrating with respect to z from $-\infty$ to ∞, the following result is obtained:

$$i\omega\mu\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}H_z(a, z')e^{-ihz'}[G_1(a, a, z-z')+G_2(a, a, z-z')]e^{-ih(z-z')}dz'dz=\int_{-\infty}^{\infty}G_1(a, b, z-0)e^{-ihz}dz. \tag{29}$$

[4] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (Technology Press, MIT, Cambridge, Massachusetts, 1946), p. 155.

[5] See reference 4, p. 151.

Let the fourier transforms of $G_1(x, x', z)$, $G_2(x, x', z)$, and $H_s(x, z)$ be denoted, respectively, by $g_1(x, x', h)$, $g_2(x, x', h)$, and $\psi(x, h)$. Then Eq. (29) is equivalent to

$$i\omega\mu\psi(a, h)[g_1(a, a, h)+g_2(a, a, h)]=g_1(a, b, h), \quad (30)$$

or

$$i\omega\mu\psi(a, h)=g_1(a, b, h)/[g_1(a, a, h)+g_2(a, a, h)]. \quad (31)$$

To find g_1 and g_2, the same procedure is followed, as described previously, in obtaining the fourier transforms of the free space Green's function. The only difference is the change of boundary conditions as required by Eqs. (22–24). It can be easily verified that the explicit expressions for g_1 and g_2 are given by the following equations,

$$g_1(x, x', h)=\begin{cases}\dfrac{1}{\lambda}\cosh\lambda(x'-a)e^{-\lambda(x-a)}, & x\geqslant x', \quad (32)\\[2ex] \dfrac{1}{\lambda}\cosh\lambda(x-a)e^{-\lambda(x'-a)}, & x\leqslant x', \quad (33)\end{cases}$$

and

$$g_2(x, x', h)=\begin{cases}\dfrac{\sinh\mu x'\cosh\mu(x-a)}{\mu\cosh\mu a}, & x\geqslant x', \quad (34)\\[2ex] \dfrac{\sinh\mu x\cosh\mu(x'-a)}{\mu\cosh\mu a}, & x\leqslant x', \quad (35)\end{cases}$$

where λ and μ have the same meaning as before. The reciprocal characteristic of a Green's function is easily recognized in Eqs. (32–35). It is now a simple matter to find E_1 and E_2. By taking the fourier transforms of Eqs. (25–26), which are

$$\phi_1(x, h)=g_1(x, b, h)-i\omega\mu\psi(a, h)g_1(x, a, h)$$
$$=g_1(x, b, h)$$
$$-\frac{g_1(a, b, h)}{g_1(a, a, h)+g_2(a, a, h)}g_1(x, a, h), \quad (36)$$

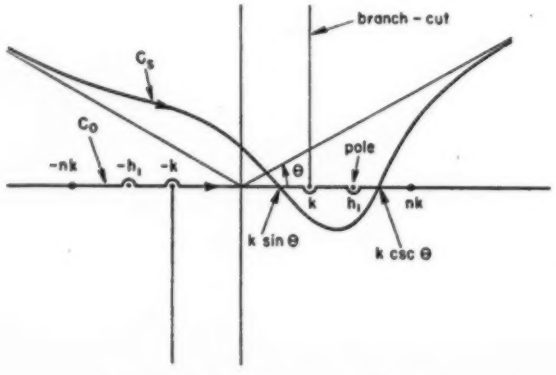


FIG. 2. Path of integration in h-plane.

$$\phi_2(x, h)=i\omega\mu\psi(a, h)g_2(x, a, h)$$
$$=\frac{g_1(a, b, h)}{g_1(a, a, h)+g_2(a, a, h)}g_2(x, a, h); \quad (37)$$

by using the appropriate forms of g_1 and g_2 given by Eqs. (32–35) as applied to different ranges of x and x', and by substituting these values into Eqs. (36–37), the same expressions for ϕ_1 and ϕ_2, as were derived by the first method, and hence the corresponding expressions for E_1 and E_2, are obtained. It should be emphasized that, in spite of the rather lengthy derivation involved in this method, an expression like Eq. (25) is very useful in finding the principal value of $E_1(x, z)$ for a certain range of x and z. In addition, a knowledge of the explicit expression for G_2 is helpful in understanding the wave propagation phenomena in the dielectric slab. In particular, as will be shown later, the normal modes associated with the eigenfunction expansion of G_2 are closely related to the residue waves due to the poles of the resultant contour integrals for E_1 or E_2.

To determine the explicit expression for G_1, Eq. (32) can be written in the form

$$g_1(x, x', h)=\frac{1}{2\lambda}\{e^{-\lambda(x-x')}+e^{-\lambda[(x-a)+(x'-a)]}\}, \quad x>x', \quad (38)$$

and similarly for Eq. (33). It can readily be seen that the right side of Eq. (38) is simply the sum of two fourier transforms of the free space Green's function, that is

$$g_1(x, x', h)=g(x-x', h)+g(x+x'-2a, h). \quad (39)$$

The inverse transform of Eq. (39), therefore, gives

$$G_1(x, x', z-z')=(i/4)[H_0^{(1)}(kr_1)+H_0^{(2)}(kr_2)], \quad (40)$$

where

$$r_1=[(x-x')^2+(z-z')^2]^{\frac{1}{2}},$$
$$r_2=[(x+x'-2a)^2+(z-z')^2]^{\frac{1}{2}}.$$

Equation (40) holds for both $x\gtrless x'$, although the derivation is carried out for $x\geqslant x'$ only.

There is no closed form for G_2. It is possible, however, to expand G_2 into a series of eigenfunctions satisfying the homogeneous wave equation and the boundary conditions specified by Eqs. (23–24). Using the method outlined, for example, in Morse and Feshbach,[6] it is found that

$$G_2(x, x', z-z')=-\frac{i}{a}\sum_{s=0}^{\infty}\frac{\sin k_s x'\sin k_s x e^{i(n^2k^2-k_s^2)^{\frac{1}{2}}|z-z'|}}{(n^2k^2-k_s^2)^{\frac{1}{2}}}, \quad (41)$$

where

$$k_s a=(2s+1)\pi/2, \quad s=0, 1, \cdots.$$

It is, perhaps, worth mentioning that the function G_2 can be interpreted as the Green's function pertaining to an idealized parallel plate dielectric wave guide, the

[6] See reference 4, p. 196.

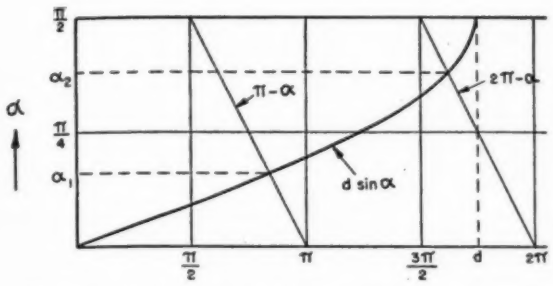


FIG. 3. Graphic solution for the equation $d\sin\alpha = n\pi - \alpha$, $n = 1, 2$, etc.

bottom surface of which is formed by a perfectly conducting plate, and the top surface by a "magnetic wall." This idealized wave guide will not support any propagating mode if k_z is larger than nk, or

$$a < \lambda_{air}/4(\epsilon)^{\frac{1}{2}}, \tag{42}$$

where ϵ, equal to n^2, is the dielectric constant of the guide. The critical depth of this idealized dielectric guide will be compared later with the actual cut-off frequency of the slab discussed in the next section.

III. CONDITION FOR THE EXISTENCE OF THE POLES

Consider the integral, Eq. (13), which has been derived by two different but equivalent methods. Since the integrand is an even function of μ for a finite value of a, the only branch points in the complex h-plane are located at $h = \pm k$ as $\lambda^2 = h^2 - k^2$. It will be seen, immediately, that if the integrand has one or more poles, they will lie along the real axis between k and nk or $-k$ and $-nk$. Both k and n are assumed to be real in this problem. The appropriate path of integration, C_0, is shown in Fig. 2 with the indentations around the poles and branch points properly indicated.

The exact locations of the poles are determined by putting

$$\mu + \lambda \tanh\mu a = 0. \tag{43}$$

Since

$$\lambda^2 - \mu^2 = (n^2 - 1)k^2, \tag{44}$$

because of Eq. (9), Eq. (43) can be rearranged into the following form, with a change of variable, $\mu a = iw$:

$$\tan w = -w/(d^2 - w^2)^{\frac{1}{2}}, \tag{45}$$

where

$$d^2 = (n^2 - 1)k^2a^2. \tag{46}$$

Following Whitmer,[2] it can be proved that Eq. (45) has no complex root of w which will satisfy the radiation condition. The only possible roots are therefore real and less than d. In discussing the numerical values of these roots, the results are considerably simplified if another change of variable is made. Thus, let

$$w = d\sin\alpha, \quad 0 \leqslant \alpha \leqslant \pi/2, \tag{47}$$

then Eq. (45) reduces to

$$\tan(d\sin\alpha) = -\tan\alpha, \tag{48}$$

or[7]

$$d\sin\alpha = n\pi - \alpha, \quad n = 1, 2, \cdots. \tag{49}$$

Because of the specified range of α, the negative integers as well as $n = 0$ are excluded. A graphical solution of Eq. (49) is sketched in Fig. 3 showing that the equation has no solution when $d < \pi/2$, or

$$a < \lambda_{air}/4(\epsilon - 1)^{\frac{1}{2}}. \tag{50}$$

When ϵ is large, the criterion is almost the same as Eq. (42), which was derived for the idealized dielectric guide described previously. If d approaches infinity, all the roots of Eq. (49) approach zero. In the complex h-plane, the poles will then form two aggregations at $h = \pm nk$ which become two essential singularities of the integrand. In other words, if Region II is a semi-infinite dielectric, the points $h = \pm nk$ become branch points in the complex h-plane in addition to the points $h = \pm k$. There will be no more isolated poles for this case. The degenerate case can be obtained directly from Eqs. (13) or (14) by putting $(x - a) = x'$, and $(b - a) = s$, and letting $\tanh\mu a \to -1$. Omitting the prime of x, we have

$$E_1(x, z) = \int_{-\infty}^{\infty} 1/4\pi\lambda \left[e^{\lambda s} - \frac{\mu + \lambda}{\mu - \lambda} e^{-\lambda s} \right] e^{-\lambda x + ihz} dh. \tag{51}$$

The path of integration denoted by C_0 is shown in Fig. 4, where two additional branch cuts have been drawn.

IV. EVALUATION OF THE CONTOUR INTEGRALS

There seem to be at least two approaches to the evaluation of the various contour integrals. Consider, for example, the integral Eq. (13) and assume there are a finite number of poles. The first method consists of

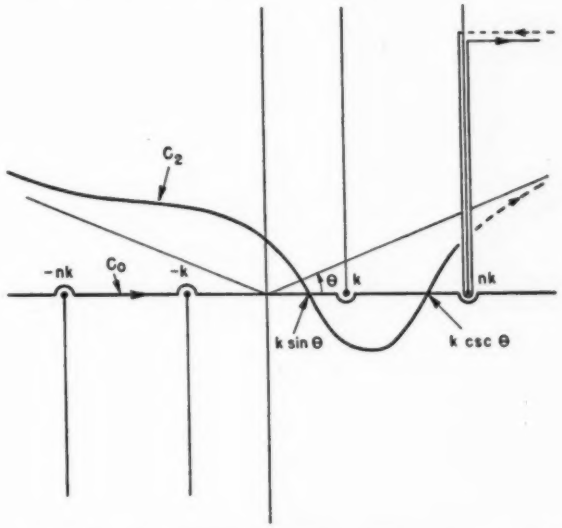


FIG. 4. Path of integration in h-plane when region II is a semi-infinite dielectric.

[7] The integer n should not be confused with the same notation used for the index of refraction.

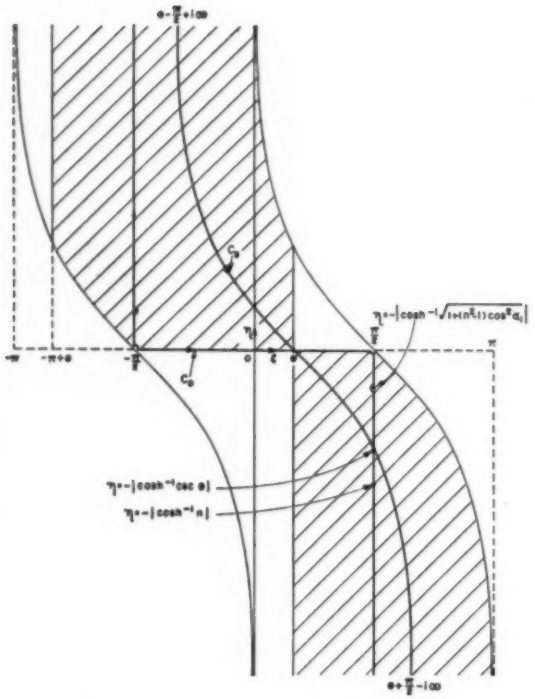


FIG. 5. Path of integration in τ-plane.

deforming the contour so that the resultant integration consists of a branch-cut integration and residue integrations around the poles. In the case of Eq. (51), where there is no pole, the deformation results in two branch-cut integrations. Although the residue integration is very easy to perform, the branch-cut integration is usually very difficult. The second approach is to employ the so-called saddle-point method of integration which deforms the path of integration in such a way that the principal part of the integral can be evaluated along this deformed path. This method is, of course, applicable only when certain parameters involved in the integrand are relatively large. For a comprehensive discussion of the application of this method to similar integrals, the reader is referred to the recent work of Ott[8] and of Cerrillo.[9] In this paper, the saddle-point of integration will be used almost exclusively to investigate the far zone electromagnetic field. It will be demonstrated that the principal part of the radiation field obtained by the saddle-point method of integration, except for certain special cases, can also be derived from the second formulation, using the fourier transform of the tangential H-field. The second derivation is, in general, much simpler, but it fails to provide the complete solution when the point of observation is near the interface.

It was stated in the introduction that the main interest of this problem was in finding the change of the radiation field which results from the presence of the dielectric slab on the conducting plane. The point of observation (x, z) is, therefore, always assumed to be at a great distance from the line source. It may, however, be very near the slab.

To apply the saddle-point integration, it is convenient to transform the integral into a complex τ-plane defined as follows:

$$h=k\sin\tau, \quad \tau=\zeta+i\eta. \tag{52}$$

If a cylindrical coordinate system is used such that

$$x-a=r\cos\theta, \quad z=r\sin\theta, \tag{53}$$

the integral defined by Eq. (13) can be written in the following form

$$E_1(r,\theta)=\int_{C_0}\frac{M(\tau)}{N(\tau)}e^{ir\cos(\tau-\theta)}d\tau, \tag{54}$$

where

$$M(\tau)=\frac{1}{2\pi}\{(n^2-\sin^2\tau)^{\frac{1}{2}}\sin[k(b-a)\cos\tau] + \cos[k(b-a)\cos\tau]\cdot\tan[ka(n^2-\sin^2\tau)^{\frac{1}{2}}]\}, \tag{55}$$

$$N(\tau)=(n^2-\sin^2\tau)^{\frac{1}{2}}-i\cos\tau\tan[ka(n^2-\sin^2\tau)^{\frac{1}{2}}]. \tag{56}$$

Since τ is a multiple-valued function of h, the domain of integration in the τ-plane corresponding to the h-plane shown in Fig. 2 is a curved strip bound by two curved lines corresponding to the two branch cuts in the h-plane. This is shown in Fig. 5 with the transformed path of integration denoted by C_0 as before. It can easily be verified that the point $h=nk$ is now transformed to $\tau=\frac{1}{2}\pi-i|\cosh^{-1}n|$, and a pole in the h-plane along the positive real axis corresponding to the root α_1 of the transcendental equation expressed in Eq. (49) is transformed into $\tau=\frac{1}{2}\pi-i|\cosh^{-1}[1+(n^2-1)\cos^2\alpha_1]^{\frac{1}{2}}|$. The saddle point of the exponential term occurring in Eq. (54), is found by setting

$$d\cos(\tau-\theta)/d\tau=0, \tag{57}$$

which gives $\tau=\theta$. The contour of stationary phase which passes through the saddle point is therefore given by

$$\cos(\zeta-\theta)\cosh\eta=1, \tag{58}$$

which is denoted by C_s in Fig. 5. The hatched areas designate that part of the τ-plane where the real part of $ir\cos(\tau-\theta)$ is negative. The corresponding contour in the h-plane, also denoted by C_s, is sketched in Fig. 2. The main task now is to deform the original path C_0 into C_s and to perform the integration along C_s. Because of the possible existence of the poles, depending on the thickness of the slab, two separate cases must be considered.

A. Case I

$$a<\lambda_{\text{air}}/4(\epsilon-1)^{\frac{1}{2}}.$$

When the condition defined by Eq. (50) is fulfilled, the integrand of Eq. (54) has no pole. For a large value

[8] H. Ott, Ann. Physik 41, 443 (1942).
[9] M. Cerrillo, *Transient Phenomena in Waveguides*, Tech. Rpt. No. 33 (Research Laboratory of Electronics, MIT, Cambridge, Massachusetts, 1948).

of r, the integral can be evaluated along the path C_0 by the classical saddle-point method.[10] The leading term of that integration gives

$$E_1(r,\theta)=M(\theta)/N(\theta)[2\pi/kr]^{\frac{1}{2}}e^{i(kr-\pi/4)}$$
$$=1/(2\pi kr)^{\frac{1}{2}}e^{i(kr+\pi/4)}\cdot f(\theta), \quad (59)$$

where

$$f(\theta)=-2\pi i[M(\theta)/N(\theta)]$$
$$=\cos[k(b-a)\cos\theta]$$
$$-\frac{(n^2-\sin^2\theta)^{\frac{1}{2}}e^{ik(b-a)\cos\theta}}{(n^2-\sin^2\theta)^{\frac{1}{2}}-i\cos\theta\tan[ka(n^2-\sin^2\theta)^{\frac{1}{2}}]}. \quad (60)$$

Equation (60) can also be written in the form

$$f(\theta)=\tfrac{1}{2}[e^{-ik(b-a)\cos\theta}+Re^{ik(b-a)\cos\theta}], \quad (61)$$

where

$$R=-\frac{(n^2-\sin^2\theta)^{\frac{1}{2}}+i\cos\theta\tan[ka(n^2-\sin^2\theta)^{\frac{1}{2}}]}{(n^2-\sin^2\theta)^{\frac{1}{2}}-i\cos\theta\tan[ka(n^2-\sin^2\theta)^{\frac{1}{2}}]}. \quad (62)$$

Equation (62) is the expression for the reflection coefficient of a horizontally polarized plane wave incident upon the slab at an angle of incidence equal to θ. Equation (61) can, therefore, be obtained directly, except for the amplitude factor, from a simple consideration of geometrical optics. The construction of the direct and reflected rays is shown in Fig. 6.

It will now be demonstrated that the expression of $E_1(r,\theta)$ as given by Eq. (59), which has been derived by applying the saddle-point method to the resultant contour integral, can be obtained directly from the second formulation, which was based upon the integration of the tangential H-field. Consider Eq. (25). It is obvious that the first term at the right side of that expression can be interpreted as a contribution to $E_1(x,z)$ due to the direct radiation from the line source and its reflection from a "magnetic wall." When the point of observation is far away from the source, it is always permissible to use the asymptotic expression of $G_1(x,b,z-0)$, which is given by

$$G_1(x,b,z-0)\doteq 1/(2\pi kr)^{\frac{1}{2}}e^{i(kr+\pi/4)}\cdot\cos[k(b-a)\cos\theta]. \quad (63)$$

Equation (63) is obtained by using the asymptotic expressions for the two hankel functions contained in Eq. (40), and by putting $r_1=r-(b-a)\cos\theta$ and $r_2=r+(b-a)\cos\theta$. In regard to the second term of Eq. (25), it can be seen, first, that

$$G_1(x,a,z-z')=(i/2)H_0^{(1)}(kr'), \quad (64)$$

where

$$r'=[(x-a)^2+(z-z')^2]^{\frac{1}{2}}. \quad 65)$$

For a fixed pair of values of x and z, the minimum value of r' occurs at $z'=z$. In order to use the asymptotic expression of Eq. (64) for all values of z', the following condition must be satisfied:

$$k(x-a)\gg 1. \quad (66)$$

In other words, the point of observation must be far away from the interface. In addition, if it is assumed that the integral of H_z contributes most when z' is in the neighborhood of the origin, so that kr' may be replaced by $k[r-z'\sin\theta]$,

$$G_1(x,a,z-z')\doteq 1/(2\pi kr)^{\frac{1}{2}}e^{i(kr+\pi/4)}e^{-ikz'\sin\theta}. \quad (67)$$

By substituting Eqs. (63) and (67) into Eq. (25), it is found that the resultant expression of $E_1(r,\theta)$ is the same as that given in Eq. (59), since

$$\int_{-\infty}^{\infty}H_z(a,z')e^{-ikz'\sin\theta}dz'=\psi(a,k\sin\theta), \quad (68)$$

in which $\psi(a,k\sin\theta)$ is the fourier transform of $H_z(a,z)$ defined at $h=k\sin\theta$. From Eqs. (31–35),

$$i\omega\mu\psi(a,k\sin\theta)$$
$$=\frac{(n^2-\sin^2\theta)^{\frac{1}{2}}e^{ik(b-a)\cos\theta}}{(n^2-\sin^2\theta)^{\frac{1}{2}}-i\cos\theta\tan[ka(n^2-\sin^2\theta)^{\frac{1}{2}}]}. \quad (69)$$

It should be noted that in deriving the expression of Eq. (59), using the second formulation, the poles of the integrand are not involved as was discussed previously. The limitation of the formula implied in the assumptions, however, cannot be easily clarified until a careful study of the contour integral is made. To illustrate this, consider a point of observation (x,z) which is very near the interface so that $k(x-a)$ is a small quantity. Then for the range of z', which is very close to z, r' becomes a small quantity. Unless $H_z(a,z')$ is small compared with $H_z(a,0)$, it is not permissible to neglect the contribution due to the integration along the interface in the neighborhood of z. In other words, if the slab supports a propagating mode, and if the point of observation is close to the interface, the formula expressed in Eq. (59), in general, no longer represents the main field. This will be explained further in the following section.

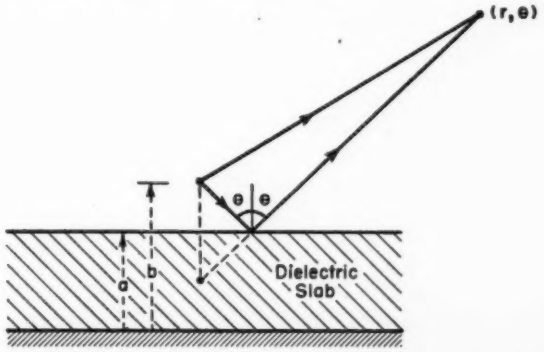


FIG. 6. Construction of direct and reflected rays based upon geometrical optics.

[10] See, for example, A. Sommerfeld, *Partial Differential Equations in Physics* (Academic Press, Inc., New York, New York), p. 98.

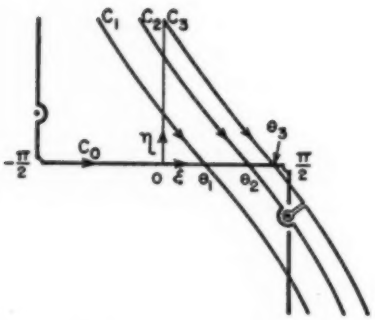


FIG. 7. The path of integration for different values of θ.

B. Case II

$$a \gtrsim \lambda_{air}/4(\epsilon-1)^{\frac{1}{2}}.$$

If the thickness of the slab is greater than the critical value, the integrand of Eq. (54) has at least two poles. For simplicity, it is assumed that

$$\lambda_{air}/4(\epsilon-1)^{\frac{1}{2}} < a < 3\lambda_{air}/4(\epsilon-1)^{\frac{1}{2}}.$$

In this case, the integrand has only two poles. The exact location of one of the poles in the τ-plane is

$$\tau_1 = \pi/2 - i|\cosh^{-1}[1+(n^2-1)\cos^2\alpha_1]^{\frac{1}{2}}|. \quad (70)$$

The other pole is at $-\tau_1$. If the contour of stationary phase is constructed as before, the path of integration, depending upon the value of θ, may run very near the pole τ_1. In Fig. 7, three curves have been drawn to show the differences. The subdivisions to be considered are: $\csc\theta$ greater than, equal to, or less than $[1+(n^2-1)\cos^2\alpha_1]^{\frac{1}{2}}$.

$$(1) \quad \csc\theta > [1+(n^2-1)\cos^2\alpha_1]^{\frac{1}{2}}.$$

In this case, the path of integration intersects the vertical axis, $\zeta=\pi/2$, at a point below the pole, as shown by C_1 in Fig. 7. Assuming that the pole is not too near the real axis, the saddle point $\tau=\theta$ and the pole are quite far apart. The integral can then be computed as in Case I, and the formula in Eq. (59) is still valid even if there is a propagating mode in the slab.

$$(2) \quad \csc\theta = [1+(n^2-1)\cos^2\alpha_1]^{\frac{1}{2}}.$$

When the path of integration intercepts the pole, an indentation must be drawn to avoid the pole, as shown by C_2 in Fig. 7. The contribution to the integral along the semicircle is equal to $i\pi R$, where R denotes the residue of the integrand which is given by

$$R = \frac{M(\tau_1)e^{ikr\cos(\tau_1-\theta)}}{dN(\tau_1)/d\tau_1}. \quad (71)$$

The remainder of the integration along the main path is still represented by Eq. (59). Since

$$\cos(\tau_1-\theta) = \cos(\tfrac{1}{2}\pi-\theta-i\eta_1)$$
$$= \sin\theta\cosh\eta_1 + i\cos\theta\sinh\eta_1, \quad (72)$$

where

$$\cosh\eta_1 = [1+(n^2-1)\cos^2\alpha_1]^{\frac{1}{2}}, \sinh\eta_1 = [(n^2-1)\cos^2\alpha_1]^{\frac{1}{2}},$$

it follows that

$$e^{ikr\cos(\tau_1-\theta)} = e^{-k(x-a)\sinh\eta_1 + ikz\cosh\eta_1}. \quad (73)$$

Therefore, because of the damping factor, the residue wave is negligible compared to the space wave as defined by Eq. (59), if $k(x-a) \gg 1/\sinh\eta_1$. (74)

$$(3) \quad \csc\theta > [1+(n^2-1)\cos^2\alpha_1]^{\frac{1}{2}}.$$

When the path of integration is deformed to C_3, as shown in Fig. 7, a full residue term, $2\pi iR$, due to the presence of the pole, must be included in the integration. The residue wave is again negligible if Eq. (74) holds true. It will be shown later that when θ is near $\pi/2$, the leading term of the saddle-point integration can be obtained directly by evaluating the branch-cut integration along the banks of the cut in the first quadrant of the h-plane, with the aid of an appropriate change of variable.

In the above discussion of the integration along the main path, it was implied that Eq. (59) is always the leading term of the saddle-point integration provided the saddle-point and the pole are not too near each other. This provision is no longer true if the pole is very near the real axis, and the saddle-point is very near the vertical axis $\zeta=\pi/2$. Physically, this corresponds to the case where the thickness of the slab is just equal to or slightly greater than its critical value, and the point of observation is very near the interface. The classical saddle-point method then fails to give an accurate answer. For this special case, the integral can be evaluated by a modified saddle-point method, as described by Cerrillo.[11] It is, however, simpler in this problem to evaluate the branch-cut integration directly. The modified saddle-point method of integration would be more valuable if the pole were on the real axis between $\zeta=0$ and $\pi/2$. Consider the integral as written in the form

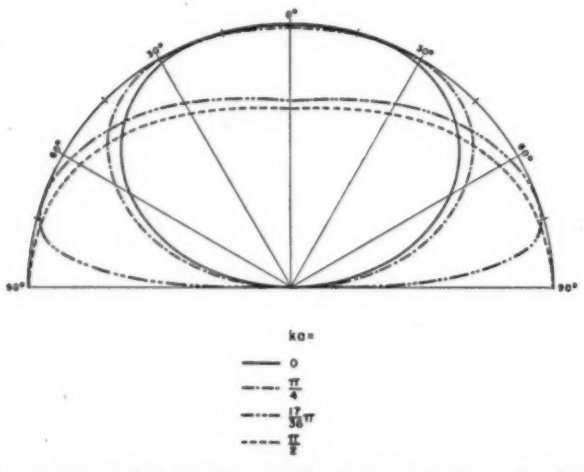


FIG. 8. The radiation pattern of the space wave, $kb=\pi/2$, $n^2=2$.

[11] See reference 9, pp. 168–169.

of Eq. (14). Let a change of variable be made such that

$$h=k+(it/z). \tag{75}$$

In terms of the new variable t, the following approximations are permissible for large values of z:

$$\lambda^2=(2ik-t/z)t/z\doteq 2ik(t/z), \tag{76}$$

$$\mu^2=-(n^2-1)k^2+(2ikt/z)-t^2/z^2\doteq-(n^2-1)k^2. \tag{77}$$

The reflection coefficient, which contains the fraction of the hyperbolic tangent function, is approximately equal to unity unless the slab is operating at its critical thickness, in which case it is equal to -1. Thus, the branch cut integration can be written as

$$E_1(x,z)=(ie^{ikz}/4\pi z)\int_{\infty}^{(+0)} e^{-t}[2ikt/z]^{-\frac{1}{2}} \times[e^{-(x-b)(2ikt/z)^{\frac{1}{2}}}\mp e^{-(x+b-2a)(2ikt/z)^{\frac{1}{2}}}]dt. \tag{78}$$

The leading term of Eq. (78) given by

$$E_1(x,z)_{\text{branch}}=\begin{cases}\dfrac{k(b-a)(x-a)}{z}\cdot\dfrac{1}{(2\pi kz)^{\frac{1}{2}}}e^{i[kz+(\pi/4)]}, & \\ & a\neq\dfrac{\lambda_{\text{air}}}{4(\epsilon-1)^{\frac{1}{2}}}, \quad (79)\\ \dfrac{1}{(2\pi kz)^{\frac{1}{2}}}e^{i[kz+(\pi/4)]}, & a=\dfrac{\lambda_{\text{air}}}{4(\epsilon-1)^{\frac{1}{2}}}, \quad (80)\end{cases}$$

is obtained by expanding the exponential functions containing x into a series and making use of the contour representation of the gamma-function.[12]

To conclude the discussion on the slab problem, the result can be summarized as follows. When the slab is operating below its critical thickness, the radiation field, which may be identified as the space wave, is given accurately by Eq. (59) for all values of θ. If the slab is thick enough to support at least one propagating mode, a surface wave exists in the neighborhood of the interface. The surface wave attenuates rapidly as the point of observation is moved away from the surface. In the space region, the principal field is still adequately represented by Eq. (59). The absolute magnitude of the function $f(\theta)$, properly normalized, is plotted in Figs. 8 and 9 for several different slab thicknesses.

V. REFLECTION FROM A SEMI-INFINITE DIELECTRIC MEDIUM

It was mentioned previously that for the extreme case when $a\to\infty$, all the roots of Eq. (49) approach zero and that the points $h=\pm nk$ become two additional branch points in the complex h-plane. When the point of observation is far away from the source, the saddle-point method of integration can still be applied to evaluate Eq. (51). However, attention must be given to the presence of the additional branch cuts. When the complex variable τ is introduced, as before, Eq. (51) can be written in the following form:

$$E_1(r,\theta)=\int_{C_0}F(\tau)e^{ikr\cos(\tau-\theta)}d\tau, \tag{81}$$

where

$$F(\tau)=\frac{i}{4\pi}\left[e^{-ik\cos\tau}-\frac{(n^2-\sin^2\tau)^{\frac{1}{2}}-\cos\tau}{(n^2-\sin^2\tau)^{\frac{1}{2}}+\cos\tau}e^{ik\cos\tau}\right], \tag{82}$$

and

$$r\cos\theta=x,\ r\sin\theta=z.$$

Portions of some of the typical paths of integration passing the saddle point $\tau=\theta$ are shown in Fig. 10. When $\sin\theta$ is less than $1/n$, the branch cut has no interference upon the path of integration, which is denoted by C_1. The radiation field is then evaluated by deforming C_0 into C_1, and the integral is computed in the usual

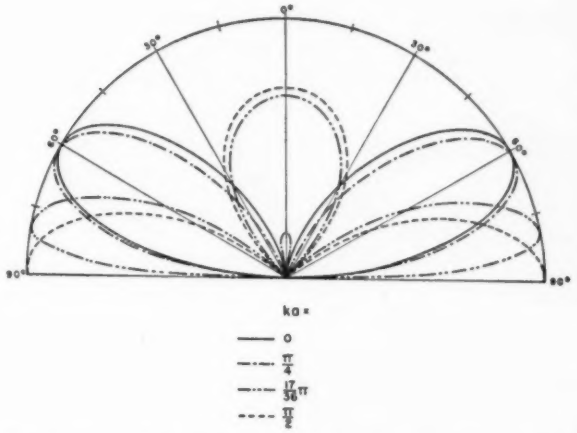


FIG. 9. The radiation pattern of the space wave, $kb=\pi$, $n^2=2$.

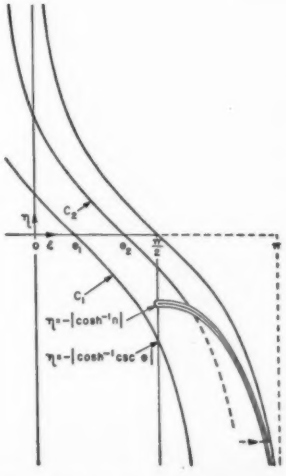


FIG. 10. Path of integration in τ-plane when region II is a semi-infinite dielectric.

[12] E. T. Whittaker and G. N. Watson, *Modern Analysis* (Cambridge University Press, The Macmillan Company, New York, 1943), p. 245.

way yielding

$$E_1(r, \theta) = F(\theta)\left(\frac{2\pi}{kr}\right)^{\frac{1}{2}} e^{i[kr-(\pi/4)]}. \quad (83)$$

If the angle of observation is such that $\sin\theta$ is greater than $1/n$, the path of integration will be intercepted by a branch cut. To evaluate the integral in this case, a branch-cut integration must be included. By following Ott's method,[8] a complete contour can be constructed, as shown by C_2 in Fig. 10, where the dotted line denotes the part of the path which is traveled on the bottom sheet of the two-sheet Rieman surface in the curved strip. The corresponding path of integration in the h-plane is shown in Fig. 4. The branch-cut integration, which is now part of the total integration, can be performed in the complex h-plane, as was described previously. The appropriate change of variable for this case is

$$h = nk + il/z. \quad (84)$$

The leading term of the branch-cut integration gives

$$E_1(x, z)_{\text{branch}} = \tfrac{1}{4}\pi[n^2/(n^2-1)][2i/\pi nkz]^{\frac{1}{2}} \times e^{[-(n^2-1)^{\frac{1}{2}}k(x+a)+inkz]}. \quad (85)$$

Because of the exponential damping factor, and also the three-halves power dependence on $1/z$, Eq. (85) is always negligible compared to Eq. (83), except at $\theta = \pi/2$, where Eq. (83) vanishes but Eq. (85) does not. The contribution due to $E_1(x, z)_{\text{branch}}$, however, is of second-order effect.

It is obvious from the general formulation of this problem that the treatment can be extended to the case where the source is imbedded in the slab. The modified problem will then be very similar to the parallel plate dielectric wave guide problem investigated by Whitmer.[2] By using the image theorem, the resultant expression for E_y can be obtained directly from his article. Evaluation of these integrals can be carried out in exactly the same manner as described in this paper for all values of θ. Finally, it should be mentioned that the problem discussed here is also related to the dielectric-filled channel guide antenna investigated by Rotman.[13] Exact treatment of the channel guide antenna is, however, more difficult because of the intricate boundary conditions. The approximate expression for the cut-off frequency of the dielectric-filled channel guide derived by Rotman is, in fact, the exact cut-off frequency for the present problem.

ACKNOWLEDGMENT

The author wishes to thank members of the Aircraft Radio Systems Laboratory who have participated in many valuable discussions of this work. In particular, he is indebted to Dr. C. Flammer for checking some important formulas given in Cerrillo's work and for many valuable discussions concerning the contour integrals derived in this paper; to Dr. W. S. Lucke who suggested the second formulation of the problem, and to Mr. A. S. Dunbar who initiated the general investigation of this problem. The assistance of Mrs. E. A. Clarke with numerical calculations is gratefully acknowledged.

[13] W. Rotman, *The Channel Guide Antenna* (Air Force Cambridge Research Laboratory, AMC, Cambridge, Massachusetts, 1950).

RCA to Show New Electronic and Radio Techniques at 1951 Conference of Corrosion Engineers

Applications of electron microscopy and radio communications were shown by the Radio Corporation of America at the 1951 Conference of the National Association of Corrosion Engineers, held at the Statler Hotel, New York City, March 13 to 16.

The new table model electron microscope, which employs magnetic lenses requiring no stabilization circuits and controls, was shown in operation. The exhibit featured the usefulness of the instrument in the study of thin films, particle size, and analysis of corrosion products. It is capable of providing direct magnifications to 6000 and photo magnifications up to 40,000 diameters.

Two-way mobile radio equipment for maintaining communications with moving vehicles and with men stationed at remote points was also on display, and the effectiveness of microwave relay systems for remote supervision, automatic operation, and telemetering purposes was highlighted at the RCA exhibit.

Formulas for the Determination of Residual Stress in Wires by the Layer Removal Method

W. T. READ, JR.
Bell Telephone Laboratories, Inc., Murray Hill, New Jersey
(Received October 2, 1950)

The distribution of residual axial stress in a beam or wire of circular cross section is derived as a function of the moment required to straighten the wire after removal of successive layers of material. Application of the formulas involves two graphical differentiations and integrations of experimental curves.

IN a recent paper[1] by R. G. Treuting and the author, a method is presented for determining the residual stress in metal sheets by removing successive layers of material from the surface of a test specimen and measuring the moment required to straighten the deformed strip. In some cases, the moment can be calculated from the curvature. This method is one of a number which depend on unbalancing the internal stress distribution so that the specimen deforms. In the present paper a formula is derived for the residual axial stress in a beam or wire of circular cross section. It is assumed that the axial stress depends only on the radial coordinate r. Successive layers of uniform thickness are removed parallel to the x axis, Fig. 1.

To obtain the axial stress σ as a function of $r=(x^2+y^2)^{\frac{1}{2}}$, we begin by considering that material has been removed so that the upper, flat, part of the surface is at $y=y_1$. Let the specimen be prevented from deforming by application of constraints at the ends. Away from the ends the stress distribution will be the same as in the original, provided the process of removing material has not introduced additional stresses. A technique of layer removal which is believed to satisfy this requirement is given in reference 1.

On the cross section shown in Fig. 1 there will be a net force

$$F(y_1)=-2\int_{y_1}^{R}dy\int_{0}^{(R^2-y^2)^{\frac{1}{2}}}\sigma(r)dx \qquad (1)$$

and a net moment about the neutral axis

$$M(y_1)=-2\int_{y_1}^{R}[y-n(y_1)]dy\int_{0}^{(R^2-y^2)^{\frac{1}{2}}}\sigma(r)dx, \qquad (2)$$

where the coordinate $n=n(y_1)$ of the neutral axis is given by

$$A(y_1)n(y_1)=\int_{-R}^{y_1}ydA=-\tfrac{2}{3}[R^2-y_1^2]^{\frac{3}{2}}, \qquad (3)$$

where

$$A(y_1)=2\int_{-R}^{y_1}(R^2-y^2)^{\frac{1}{2}}dy$$

$$=\left[y_1(R^2-y_1^2)^{\frac{1}{2}}+R^2\left(\frac{\pi}{2}+\sin^{-1}\frac{y_1}{R}\right)\right]. \qquad (4)$$

By equilibrium, the force F and moment M due to the residual stress σ are equal to the external force and moment required to prevent deformation. Since the moment is much more easily measured than the force, we shall work with Eq. (2), which relates the stress $\sigma(r)$ to the moment M. If the material is linear and elastically homogeneous, M can be calculated from the curvature of the specimen.

Some manipulation is required in order to obtain $\sigma(r)$ explicitly from (2). This is carried out as follows.

Differentiating (1) gives

$$\frac{dF}{dy_1}(y_1)=2\int_{0}^{(R^2-y_1^2)^{\frac{1}{2}}}\sigma(r)dx. \qquad (5)$$

Two steps are now required to determine $\sigma(r)$ as a function of the measured $M(y_1)$. First, $F(y_1)$ is obtained in terms of $M(y_1)$ by substituting (1) and (5) in the derivative of (2) and solving the resulting differential equation. Second, (5) is put in the form of Abel's integral equation, for which the solution is known.

To find F in terms of M we differentiate (2), obtaining

$$\frac{dM}{dy_1}=2[y_1-n(y_1)]\int_{0}^{(R^2-y_1^2)^{\frac{1}{2}}}\sigma(r)dx$$

$$+\frac{dn(y_1)}{dy_1}2\int_{y_1}^{R}dy\int_{0}^{(R^2-y_1^2)^{\frac{1}{2}}}\sigma(r)dx. \qquad (6)$$

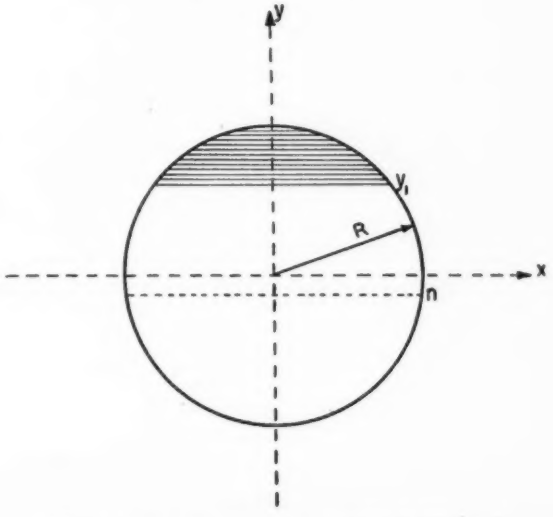


FIG. 1. Cross section of wire. Layers of material are removed from upper part.

[1] R. G. Treuting and W. T. Read, J. Appl. Phys. **22**, 130 (1951).

Substituting (1) and (5) into (6) we have the differential equation

$$\frac{dM}{dy_1}=(y_1-n)\frac{dF}{dy_1}-F\frac{dn}{dy_1}. \tag{7}$$

Using (3) to obtain

$$\frac{1}{y_1-n}\frac{dn}{dy_1}=\frac{1}{A}\frac{dA}{dy_1}, \tag{8}$$

(7) is readily integrated to give

$$F(y_1)=A(y_1)\int_0^{M(y_1)}\frac{dM}{A(y)[y-n(y)]}. \tag{9}$$

The integral in (9) is evaluated by plotting $1/A(y)[y-n(y)]$ *vs* $M(y)$ and integrating graphically from $M(R)=0$ to $M(y_1)$. The derivative of $F(y_1)$ is then obtained graphically and used in the solution of the integral equation (5), which can be solved explicitly for $\sigma(r)$, as we now show.

We first define the functions

$$\sigma_1\left(\frac{1}{r^2}\right)=\sigma(r),\quad f_1\left(\frac{1}{y_1^2}\right)=\frac{dF(y_1)}{dy_1}, \tag{10}$$

and variables

$$v=1/r^2,\quad v_1=1/y_1^2, \tag{11}$$

in terms of which (5) becomes

$$f_1(v_1)=(v_1)^{\frac{1}{2}}\int_{v_0}^{v_1}\frac{\sigma_1(v)}{v^{\frac{1}{2}}}\frac{dv}{(v_1-v)^{\frac{1}{2}}}, \tag{12}$$

where

$$v_0=\frac{1}{R^2}.$$

Defining two new functions,

$$S(v)=\frac{\sigma_1(v)}{v^{\frac{1}{2}}}$$

$$g(v_1)=\frac{f_1(v_1)}{(v_1)^{\frac{1}{2}}}, \tag{13}$$

Eq. (12) becomes

$$g(v_1)=\int_{v_0}^{v_1}\frac{S(v)dv}{(v_1-v)^{\frac{1}{2}}}, \tag{14}$$

which is Abel's integral equation and has the solution[2]

$$S(v)=\frac{1}{\pi}\frac{d}{dv}\int_{v_0}^{v}\frac{g(v_1)dv_1}{(v-v_1)^{\frac{1}{2}}}. \tag{15}$$

Changing back to the original variables and functions, (15) becomes

$$\sigma(r)=-\frac{1}{\pi}\frac{d}{dr}\int_r^R\frac{r}{y_1}\frac{1}{(y_1^2-r^2)^{\frac{1}{2}}}\frac{dF}{dy_1}(y_1)dy_1. \tag{16}$$

It will be convenient to define

$$w=\cos^{-1}r/y_1$$

$$H(w,r)=\frac{dF}{dy_1}(y_1), \tag{17}$$

where $dF(y_1)/dy_1$ is given by (9). In terms of the variable w and function H, (16) becomes

$$\sigma(r)=-\frac{1}{\pi}\frac{d}{dr}\int_0^{\cos^{-1}r/R}H(w,r)dw. \tag{18}$$

By plotting H *vs* w for various values of r, $\sigma(r)$ is obtained from (18) by graphical integration and differentiation.

Because of the two-dimensional nature of the problem, two graphical differentiations and integrations are required instead of one as in the case of the strip.

It will be convenient to use a different specimen for each depth of removal to avoid the experimental difficulty of removing successive parallel layers from a single round wire. However, this is not a drawback, since a number of identical specimens can be cut from a wire which has been drawn by a continuous and uniform process.

The author wishes to thank Dr. L. A. MacColl for valuable suggestions regarding the application of Abel's integral equation and Dr. R. G. Treuting for initiating and encouraging the analysis.

[2] E. T. Whittaker and G. N. Watson, *Modern Analysis* (Cambridge University Press, 1920), third edition, p. 229.

Directional Distribution of 1040-Kev Radiation from a High Voltage X-Ray Tube

N. GOLDSTEIN, U. FANO, AND H. O. WYCKOFF
National Bureau of Standards, Washington, D. C.
(Received September 27, 1950)

The distribution of the 1040-kev x-rays emitted in various directions by a high voltage x-ray tube has been measured by means of an indium resonance detector.

ACCORDING to Miller and Waldman,[1] the 4.4-hour radioactivity of the isomeric metastable state of indium (In^{115*}) induced by a complex x-radiation indicates the intensity of the 1040-kev component in that complex, since the activity is induced by absorption of a sharp line of the x-ray spectrum. Since there is a second absorption line at 1420 kev, this method is easily applicable only if the peak energy of the spectrum lies below this voltage. The indium method has thus far been used to analyze the radiation from a thick target without regard to the angular distribution of the radiation.

Inasmuch as very little is known experimentally,[2] or theoretically about the spectral distribution of the radiation emitted in various directions by a high voltage x-ray tube, it seemed worthwhile to apply the indium method to a study of the radiation from a 1400-kv constant potential x-ray tube.[3] This tube, which is energized by a rectifier source, delivers a much high intensity than the generator used by Miller and Waldman. Therefore, it affords a welcome opportunity to extend the application of the indium activation technique whose sensitivity is very low.

The results reported in this paper seem perfectly consistent with those reported by Miller and Waldman. However, departures of the resonance levels from the Miller-Waldman values by a few tens of kilovolts could have remained unnoticed.

Even though a relatively intense x-ray beam was used, the present experiment was subject to severe limitations. The low intensity of the measured activity required geometry conditions which were far from ideal. Data from a thin target geometry would be of greater basic interest, whereas data from a thick target is of greater practical importance. At any rate, a thick target was used to obtain adequate intensity. Under these conditions the incident electrons could undergo substantial scattering prior to x-ray emission. In addition, the cooling of the target involves the use of a substantial layer of water and other materials giving rise to absorption and Compton scattering between the source and the detector. Even though the present results serve primarily for orientation purposes, they seem worth reporting in the absence of other comparable data.

The geometrical arrangement is illustrated in Fig. 1. The indium detectors were strips of foils 9.5 inches by 1.5 inches and 0.01 inch thick. They were rolled into cylinders $\frac{3}{8}$ inch in diameter and $\frac{1}{2}$ inch long to occupy a small volume near the target. After irradiation they were unrolled and spiralled around a Victoreen 1B85 aluminum counter having a wall about 30 mg/cm² thick. The activity of the foils exposed at different angles in each run were measured alternately for several hours after irradiation. Allowance was made for the 105-minute activity due to the metastable state of In^{113} (abundance 4.2 percent), in extrapolating the curves back to the time irradiation was completed. Figure 2 shows a typical plot of the activity counts and of their extrapolation to zero-time activity. The usual irradia-

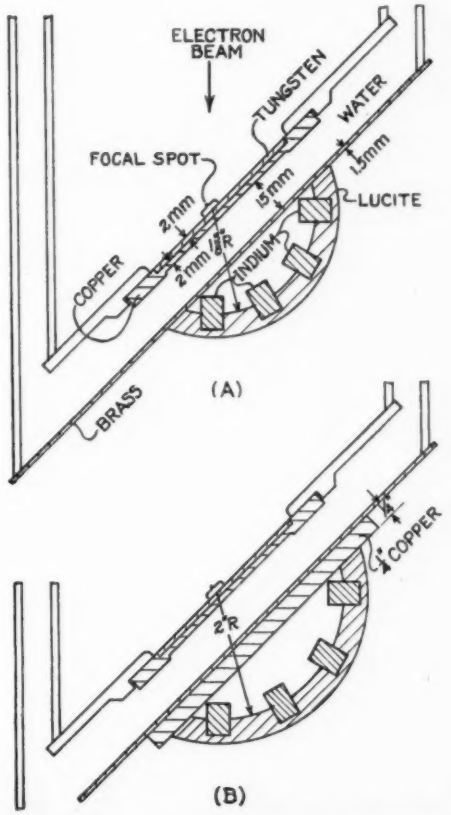


FIG. 1. Experimental arrangement without and with ¼-inch copper between target and detectors.

[1] W. C. Miller and B. Waldman, Phys. Rev. **75**, 425 (1949).
[2] The extensive experiments by Buechner, Van de Graaff, Burrill, and Sperduto, reported in the Phys. Rev. **74**, 1348 (1948) bear only on the total intensity radiated in various directions, rather than on the intensity of a particular monochromatic component.
[3] E. E. Charlton and H. L. Hubbard, Gen. Elec. Rev. **43**, 271 (1940).

TABLE I. Counts per minute corrected to 5-hr irradiation.

Run No.	No Cu absorber (Fig. 1-A)						¼-in. Cu absorber (Fig. 1-B)		
	2	7	8	1	6	3	9	4	5
Voltage (kv)	1400	1400	1400	1200	1200	1100	1100	1400	1200
Angle: 0°	455	550	500	250	270	91	117	230	130
(with 30°		500	510	235	240	61	65	224	90
respect to 60°	316	320	340		115	22	19	142	44
electron 90°	192	190	220	61	64	4	4	83	19
beam)									

tion was for five hours; otherwise, the activity data were adjusted to that time interval.

Runs were made at 1100-, 1200-, and 1400-kv applied potential with a current of 0.5 ma. As shown in the figure, the focal spot is not too small (about 8 mm diameter), and may have been wandering somewhat. Wandering by 2 mm could account for the limited reproducibility of the experimental results. The complete set of results is given in Table I and illustrated in Fig. 3. The counting rates indicated in the figure for the 30° and 60° positions have been decreased by 17 percent to correct for the smaller absorption in these directions.

The tabulation includes data taken with an additional layer of ¼-inch copper inserted between the x-ray source and the indium detectors. This experiment was made to examine to what extent the change in activity, due to attenuation in the copper layer, follows a simple law.

There doesn't seem to be any purpose in attempting to compare the results quantitatively with detailed theoretical calculation of the bremsstrahlung spectrum. The theory itself is not in a very satisfactory state, especially in this energy range which is neither nonrelativistic nor extreme relativistic.[4] The experimental results are greatly complicated by the energy loss and scattering of the electrons in the thick target. The total probability of absorption plus scattering of 1040-kev

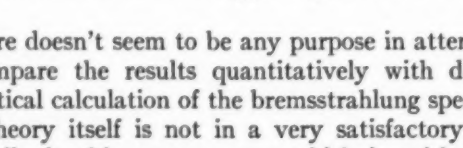


FIG. 2. Sample plot of measurements of indium radioactivity. The straight line indicates the extrapolation back to the end of the irradiation discounting the weak shorter lived activity of the In^{113} isotope.

[4] W. Heitler, *The Quantum Theory of Radiation* (Oxford University Press, Oxford, 1944), second edition, p. 161 ff.; and especially A. Sommerfeld, *Atombau und Spektrallinien* (Friedrish Vieweg & Sohn, Braunschweig, Germany, 1939), second edition, Vol. 2, Chap. 7.

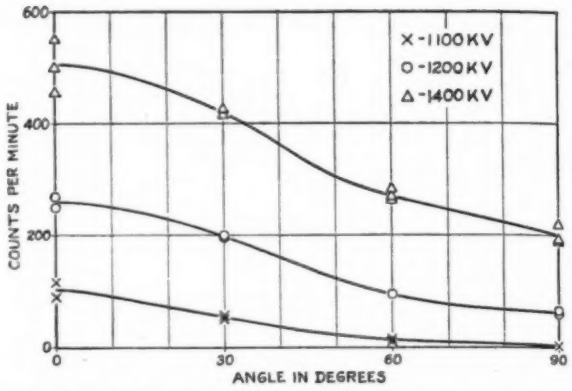


FIG. 3. Plot of the activities induced in different directions with respect to the incident beam for several different voltages applied to the tube.

x-rays in the materials between the target and the detector is approximately 40 percent.

The intensity of 1040-kev x-rays is increased by Compton scattering of the higher energy components in these materials. This increase amounts to approximately 8 percent for an applied potential of 1400 kv according to a calculation by Miss G. R. White.

All the directional intensity distributions are peaked in the direction of the incident electron beam, as one should expect, since the x-rays take up much of the momentum of the irradiating electrons. However, the distributions are much less peaked for the higher voltage runs than for the lower voltage runs. The main contribution to this greater angular spread arises from the ability of higher energy electrons to undergo a large deflection before their energy falls below the level required to activate the indium detector. A numerical calculation by Dr. W. Miller, following the method of H. W. Lewis,[5] verifies that the incident electrons become nearly isotropic after they have lost about 200 kev in the tungsten target. This estimate, which agrees with Bothe's results[6] accounts roughly for the observed effects. An additional effect arises from the fact that higher energy electrons have a greater chance of producing excess energy photons whose energy falls to 1040 kev after Compton scattering.

The data obtained with an additional copper layer between the x-ray source and the detector do not indicate that the layer causes a clear and significant change of angular distribution within the low accuracy of the experiment. The reduction of intensity caused by inserting the copper layer amounts to approximately 50 percent. Of this, about 34 percent is caused by the increased distance of the indium from the x-ray source (2 inches instead of $1\frac{5}{8}$ inches). The remaining attenuation corresponds almost exactly to the "narrow beam" attenuation of 1040-kev x-rays in ¼ inch of copper calculated according to the formula

$$\exp[-(0.51\ \text{cm}^{-1})(0.63\ \text{cm})].$$

[5] Phys. Rev. **78**, 526 (1950).
[6] Ann. Physik **6**, 44 (1949).

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 4 APRIL, 1951

Lattice Distortion Spectrum of Evaporated Gold

P. G. WILKINSON
U. S. Naval Research Laboratory, Washington, D. C.
(Received August 21, 1950)

Gold was deposited in various pressures of nitrogen from 0.001μ to 110μ at low temperatures. The electrical resistance of the gold film was observed to decrease as it warmed, enabling the distribution of lattice distortions to be computed as a function of decay energy. The distortion spectrum from $70\cdot10^{-14}$ to $130\cdot10^{-14}$ erg was found to consist of three peaks—one at $80\cdot10^{-14}$, one at $100\cdot10^{-14}$, and one at $120\cdot10^{-14}$ erg. The relative number of distortions for a particular decay energy was found to decrease as the nitrogen pressure increased. The final resistivities of the gold films were shown to be dependent on the film thickness over the range 40A to 420A in a manner which agreed with theory and gave a value of 970A for the mean free path of the conduction electrons.

INTRODUCTION

IT is well known that metallic films condensed from the vapor possess high initial electrical resistivities which decay irreversibly and approach the resistivity of the bulk metal. The initial high resistivity is due to the presence of a large number of crystal lattice distortions caused by the condensation process. These so-called combined-type distortions consist of lattice deficiencies located close to lattice excesses, and can decay by small scale diffusion through the lattice. If the metal is deposited upon a surface cooled with liquid air, the initial high resistivity can be preserved indefinitely, especially if the metal is cooled below the deposition temperature. If the substrate is heated uniformly, the rate of decay of resistance will attain maxima at certain absolute temperatures, each maximum corresponding to a definite number of distortions in the freshly formed film. Vand[1] has formulated mathematically the theory of this decay and has shown how to calculate the "decay energy" and the relative number of distortions from measurements of the rate of resistance decrease with time and temperature. It thus becomes possible to construct an energy spectrum for the distribution.

In a previous paper,[2] measurements were made on the decay of resistance of gold films deposited in high vacuum onto glass and quartz substrates. Calculations of the decay energy and the distribution of distortions revealed these to be functions of the rate of deposit and the type of substrate. Actually, the number of distortions and their characteristic decay energy are affected to a great extent by the presence of minute impurities, and residual gas molecules during the evaporation, and to a lesser extent by the rate of deposit, film thickness, type of substrate, and temperature of deposition.

Metal films formed by cathodic sputtering also show the decay of resistance phenomenon as has been observed for gold, by Pogany,[3] Perrucca,[4–8] Kramer and Zahn,[9] and for other metals by King,[10] Braunsforth,[11] and Janitzky.[12]

In the present experiments, gold was condensed at various pressures of nitrogen from 0.001μ to 110μ and the decay of electrical resistance measured. The spectrum of the distribution of distortions was then computed and found to consist of three maxima, at about $80\cdot10^{-14}$, $100\cdot10^{-14}$, and $120\cdot10^{-14}$ erg, respectively, the exact positions depending upon the conditions of preparation. The number of distortions was found to decrease as the nitrogen pressure increased.

Even after aging at room temperature, the resistivities of very thin evaporated metallic films are higher than the bulk resistivity due to a shortening of the mean free path of the conduction electrons by collision with the film boundaries.[3,9,13–17] The final resistivities of the gold films, after the distortions decayed, were found to be a function of film thickness in a manner which agreed with theory and gave 970A as the mean free path of the conduction electrons.

EXPERIMENTAL

All the evaporations were made in an all Pyrex evaporator constructed from a five-liter flask which could be readily evacuated to 10^{-6} mm pressure and to which was attached a gas purification train, storage vessels, manometers, McLeod gauges, liquid air traps, and a gas analysis system. The electrical resistance measurements were done on silver-tipped microscope slides possessing a clear area measuring 1 cm by 1½ cm, and located on the bottom of a flat brass plate attached to a glass tube which was kept filled with liquid air during an evaporation. The evaporations were made

[1] V. Vand, Proc. Phys. Soc. **55**, 222 (1943).
[2] P. G. Wilkinson and L. S. Birks, J. Appl. Phys. **20**, 1168 (1949).
[3] B. Pogany, Ann. Physik Chemie **49**, 531 (1916).
[4] E. Perrucca, Ann. Phys. Leipzig **4**, 252 (1930).
[5] E. Perrucca, Nuovo Cimento **7**, 50 (1934).
[6] E. Perrucca, Nuovo Cimento **11**, 534 (1937).
[7] E. Perrucca, Nuovo Cimento **14**, 506 (1938).
[8] E. Perrucca, Nuovo Cimento **15**, 365 (1938).
[9] J. Kramer and H. Zahn, Naturwiss. **20**, 792 (1932).
[10] R. W. King, Phys. Rev. **10**, 291 (1917).
[11] G. Braunsforth, Ann. Phys. **9**, 417 (1931).
[12] A. Janitzky, Z. Physik **31**, 277 (1925).
[13] A. C. B. Lovell, Proc. Roy. Soc. (London) **A157**, 311 (1936).
[14] E. T. S. Appleyard and A. C. B. Lovell, Proc. Roy. Soc. (London) **A158**, 718 (1937).
[15] K. Fuchs, Proc. Cambridge Phil. Soc. **34**, 100 (1938).
[16] R. A. Weale, Proc. Phys. Soc. (London) **A62**, 135 (1949).
[17] W. Planck, Physik. Z. **15**, 563 (1914).

from an electrically-heated tungsten wire cup filled with gold wire of high purity (less than 0.1 percent impurities, mostly copper) and located 7 cm below the cold condensing surface.

Specimen temperatures were read by means of a calibrated copper-constantin thermocouple and Rubicon potentiometer. Resistance measurements were made with a Leeds and Northrup Wheatstone Bridge

The evaporations at 10^{-6} and 10^{-5} mm residual gas pressure were carried out with continuous pumping at these pressures. Other evaporations were done in 0.14μ, 0.88μ, 8.5μ, 10.6μ, and 110μ of static nitrogen purified by passing tank nitrogen over copper gauze heated to 500°C and dried over phosphorus pentoxide. This gas analyzed 0.001 percent oxygen which was low enough to avoid formation and evaporation of tungsten oxides.[18]

The thickness of the gold was found by dissolving a measured area in aqua regia, evaporating to dryness, and determining the gold chloride colorimetrically by reaction with *o*-tolidine reagent.[19] The density of vacuum-evaporated gold was determined by Kahlbaum, Roth, and Siedler[20] and Kahlbaum and Sturm[21] as 18.88 g/cm³ which is slightly less than that of bulk gold (19.3 g/cm³).

THE DECAY THEORY OF LATTICE DISTORTIONS

The Resistivity of Metals

From theory and experiment,[1,2] it has been established that the electrical resistivity of most metals is the sum of three contributions:

$$\rho=\rho_T+\rho_B+\rho_i, \tag{1}$$

where ρ_T results from the scattering of conduction electrons upon the periodic thermal oscillations of the atoms in the crystal lattice; ρ_B results from a scattering of the conduction electrons on the walls of the conductor, and is important only for very thin films in which the thickness is small in comparison to the mean free path of the conduction electrons; and ρ_i results from a scattering of the conduction electrons on foreign atoms and lattice distortions. In our experiments, ρ_i was so much larger than ρ_T and ρ_B that these latter could usually be neglected.

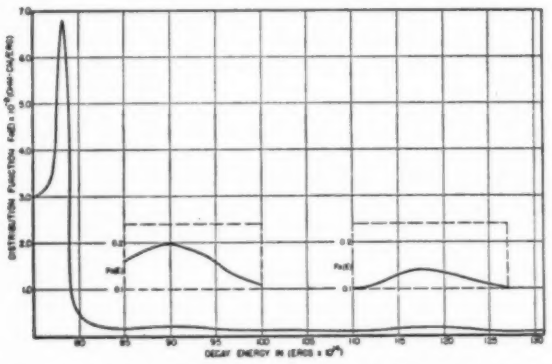


FIG. 1. The distortion spectrum of a gold film 44A thick deposited in 8.5μ of nitrogen. The inserts are sections of the curve magnified to show the shape of the higher energy portion.

Experimental Facts

The theory of the decay of lattice distortions developed by Vand[1] is based upon the experimental fact that when the temperature of a freshly formed film is raised at a constant rate above the temperature of preparation, the resistance remains constant for a time, then undergoes a succession of several rapid irreversible falls. Each fall of resistance occurs at a characteristic temperature and is followed by a period of comparatively slow resistance change. Finally, on standing at room temperature, a low stable resistance is reached which exceeds the bulk value by virtue of the small thickness, traces of impurity atoms, and undecomposed distortions.

Types of Distortions

The distortions formed in a metal by the condensation process are usually considered to be a combination of closely spaced lattice deficiencies and compressions which may neutralize one another by recombination (recombination theory) or each distortion may regroup itself or "decay" into a more ordered lattice by a small-scale diffusion process (decay theory). The decay theory of combined distortions is the more reasonable since it leads to values of the decay energy which are of the same order of magnitude as the binding energy of the metal atoms.

Assumptions of the Decay Theory

The decay theory of the combined type of distortions is based upon the following assumptions: (a) Each combined distortion has an energy content higher than the same number of atoms in an ordered lattice, and, at any given instant, a finite probability exists that it may attain a critical energy or "energy of activation" which will allow the system to pass over the potential barrier and reach a state of lowest energy. (b) Each distortion scatters the conduction electrons independently and its decay does not influence the decay of its neighbors. (c) All distortions can be so ordered according to their decay energy that in any small interval ΔE, there are $N(E)\Delta E$ distortions. (d) The total value of the contributions of the lattice distortions to the resistivity is:

$$\rho_i=\int_{-\infty}^{+\infty} r(E)N(E,t)dE, \tag{2}$$

[18] P. G. Wilkinson, J. Appl. Phys. **22**, 226 (1951).
[19] W. S. Clabaugh, J. Research Natl. Bur. Standards **36**, 119 (1946).
[20] Kahlbaum, Roth, and Siedler, Z. anorg. u. allgem. Chem. **29**, 177 (1902).
[21] G. W. A. Kahlbaum and E. Sturm, Z. anorg. u. allgem. Chemie **46**, 217 (1905).

TABLE I. Classification of distortions in gold films.

Run No.	Pressure of nitrogen (μ Hg)	Temp. of condensation (°K)	Thickness (A)	Final resistivity (μ ohm-cm 0°C)	Rate of deposit (μ g/cm²-min)	Class A $F_0(E)\cdot10^{-8}$	Class A $E\cdot10^{14}$	Class B $F_0(E)\cdot10^{-8}$	Class B $E\cdot10^{14}$	Class C $F_0(E)\cdot10^{-8}$	Class C $E\cdot10^{14}$
1	0.001	140	130	—	7.0	—	<70	48.3	106	78	119.7
2	0.01	126	49	55.4	3.1	12.8	77.6	—	—	2.4	124.2
3	0.01	154	90	—	10	—	<70	12.8	95.5	—	—
4	0.01	154	44	—	2.8	—	<70	—	—	14	124.4
5	0.14	127	226	11.08	28	—	<70	0.42	100.5	0.33	112
6	0.14	127	222	13.2	28	—	<70	1.30	101.7	0.22	113
7	0.88	131	353	8.11	27	10.6	93	0.30	103.4	0.031	123.4
8	0.88	131	420	8.81	32	—	<70	0.101	103.2	0.17	123.7
9	8.5	136	44	60.6	2.8	6.8	78.4	0.195	89.5	0.14	117
10	10.6	147	185	13.8	10	0.79	84	—	—	0.24	113.5
11	10.6	147	232	16.7	12	1.24	84.4	—	—	0.50	114.1
12	110	115	99	34.5	6	1.13	78.2	—	—	—	—
13	110	115	84	—	5	0.53	73.3	—	—	—	—

where $r(E)$ is the average contribution of one distortion per unit volume to the resistivity; and $N(E, t)$ is the number of distortions of decay energy, E, in unit volume at time, t. The determination of $N(E, 0)$, the number of distortions at the instant of film formation, ordered according to their decay energy, is the object of the theory.

By measuring the fall of electrical resistance with a uniformly rising temperature, the function $r(E)N(E, 0) = F_0(E)$, though not $N(E, 0)$ alone, may be determined. $F_0(E)$ should have properties similar to $N(E, 0)$ as $r(E)$ probably does not change from specimen to specimen.

The Mathematical Results

Without repeating Vand's mathematical development, the results may be tabulated. The decay energy, E, is given by:

$$E = ukT, \tag{3}$$

where T is the absolute temperature, k is Boltzmann's constant, and u is a function of time, t:

$$u + \log u = \log 4nft, \tag{4}$$

where n is the number of atoms which can initiate the decay of a single distortion, estimated to be about 10, and f is the maximum Debye frequency of vibrations in the perfect lattice ($3.55\cdot10^{12}$/sec for gold) and is assumed to be the same for the distorted lattice also. The distribution function may be expressed:

$$F_0(E) = -d\rho_i/dT/kU, \tag{5}$$

where

$$U = u(u+2)/u+1. \tag{6}$$

Because of the approximations involved in the solution for Eq. (5) $F_0(E)$ may be in error by $2kT/E$.

EXPERIMENTAL MEASUREMENTS OF THE DISTRIBUTION FUNCTION

In order to calculate the distribution function, it is necessary to condense the metal at a low temperature and measure the resistance as a function of time and temperature as the film warms up. From the thickness of the film, $d\rho_i/dT$ may be calculated from these experimental measurements. The time dependent functions, u and U, may be computed from Eqs. (4) and (6). It is then a simple matter to compute from Eq. (5) $F_0(E)$ as a function of the decay energy.

The Spectrum of the Distribution Function

By employing evaporating times less than 3 minutes, cooling the deposition surface 30°C lower than previously (−100°C), and working with metal films as thin as 40A, the complete distortion spectrum of evaporated gold from −130°C to 0°C was obtained. This is illustrated in Fig. 1 which shows the distribution function for a film 44A thick condensed at 136°K in 8.5 microns of nitrogen. The decay energies included were from $76\cdot10^{-14}$ erg to $130\cdot10^{-14}$ erg, which represents a temperature scale of −117°C to −15°C. Three maxima in the number of distortions per cc exist, one at $78.4\cdot10^{-14}$ erg, one at $89.5\cdot10^{-14}$ erg, and one at $117\cdot10^{-14}$ erg, with values for $F_0(E)$ of $6.8\cdot10^{8}$, $0.195\cdot10^{8}$, and $0.14\cdot10^{8}$, respectively. Since neither the absolute value of $r(E)$ nor its variation with decay energy is known, these values of $F_0(E)$ do not reveal the absolute number of distortions present at different energies. The distribution function does reveal that there are three preferred values of the decay energy, that the one of lowest energy in this case is the most populated, that the one of intermediate energy is much less populated and that the high energy is the least populated.

Classification of the Spectrum

Thirteen evaporations of gold were made in different pressures of nitrogen in order to evaluate the effect of molecular collision and gas occlusion on the distribution function. Data for these runs are summarized in Table I.

From a study of the characteristics of each distribution curve, it seemed reasonable to specify as Classes A, B, and C, those types of distortion having decay energies of $70\cdot10^{-14}$ to $90\cdot10^{-14}$ erg, $90\cdot10^{-14}$ to $110\cdot10^{-14}$ erg,

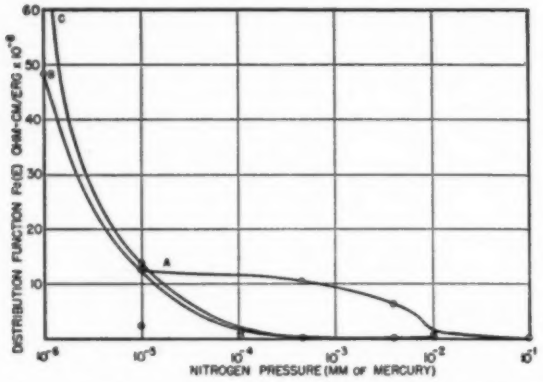


FIG. 2. The distribution function $F_0(E)$ at increasing nitrogen pressures. A, B, and C represent the types of distortions classified by their decay energies (see text).

and $110 \cdot 10^{-14}$ to $130 \cdot 10^{-14}$ erg, respectively. Vand[1] has calculated decay energies from the data of Suhrmann and Barth[22] on gold films condensed on quartz and obtained energies of $95 \cdot 10^{-14}$, $131 \cdot 10^{-14}$, and $227 \cdot 10^{-14}$ erg. The first two agree with our assignment, and the last was obtained beyond the range of our measurements. Table I gives values of the distribution function, $F_0(E)$, and the decay energy, E, for these classes. All three classes are usually not observed in any given sample although all are present in runs 7 and 9. The reason for this is not obvious but some remarks can be made. It has been shown[2] that a variation in the rate of deposit from 1 to 5 micrograms/sq cm/min causes a 20 percent increase in the decay energy for the Class C distortion. Other factors which would be expected to influence the decay energy are film thicknesses below 200A,[2] substrate temperature, and pressure of nitrogen during evaporation. In general, these same factors would influence also the distribution function. If the decay energy were shifted to slightly below $70 \cdot 10^{-14}$ erg, only the high energy side of the distribution function would be found. Actually, only the high energy side of the Class A distortion was found for Runs 3, 4, 5, 6, and 8 and the maximum was not reached.

Class B distortions were not found for Runs 2 and 4 possibly because the rates of deposit of 3.1 and 2.8 micrograms/sq cm/min were too low to form these. Also, no Class C distortion was observed for Run 3. Another possibility exists that the resolving power of the method of computation, which is of the order of kT, is not great enough to reveal these. For the Class B distortion in Runs 2 and 4, the resolving power is about $3 \cdot 10^{-14}$ erg and for the Class C distortion in Run 3 it is about $4 \cdot 10^{-14}$ erg.

The Effect of Nitrogen Pressure on the Distribution Function

In a plot of the distribution function against nitrogen pressure for each class of distortion (Fig. 2), it is noticed that $F_0(E)$ decreases rapidly at pressures of 0.001μ to 0.1μ and becomes very small for Class B and C distortions at pressures higher than 1μ. Actually from Table I, it is seen that there are no characteristic maxima at 10.6μ and 110μ for Class B and at 110μ for Class C distortions. The number of Class A distortions, however, is not a sensitive function of pressure until about 8μ-pressure is reached, and it decreases more rapidly above this pressure, becoming very small at 110μ.

Conclusions

This dependency of the distribution function on pressure is interpreted to mean that nitrogen molecules either transfer kinetic energy to a distortion by collision and cause it to decay spontaneously, or that nitrogen molecules are condensed with the metal on the collecting surface and prevent the distortion from forming. The average kinetic energy of a nitrogen molecule at 300°K is only $6.2 \cdot 10^{-14}$ erg which is about 5–10 percent of the energy required. Only very exceptional molecules (about one in 10^{11}), would have an energy great enough to be effective. Moreover, the Class A distortion, which has the lowest decay energy, is affected the least by the presence of nitrogen. It is concluded that nitrogen molecules trapped in the crystal lattice must in some manner contribute to the formation of a more ordered lattice or that they can cause a completely new type of distortion to form which contributes little to the total resistivity.

THE RESISTIVITY OF THIN GOLD FILMS

The resistivities of evaporated gold films thinner than about 5.7 micrograms/sq cm (corresponding to 30A in statistical thickness) are abnormally high due mostly to the formation of small isolated globules of the metal as revealed by electron microscope photographs.[2] As the deposit is built up, the isolated globules grow together[23] and an apparently continuous film of low resistivity is formed. Although there exists considerable surface roughness and some contact resistance between the gold particles, the resistivity of gold films can be shown experimentally to be a function of film thickness[3] owing to a shortening of the mean free path of the conduction electrons by collision with the walls of the conductor.

Experimental Results

Although the gold films described in this research were produced at low temperatures in varying pressures of nitrogen, and undoubtedly contained small amounts of occluded nitrogen, their resistivities, after the distortions decayed, showed a marked dependency upon film thickness (circles in Fig. 3). The resistivity approached the value of gold in the bulk (2.04 microhm-cm at 0°C) at film thicknesses in excess of 400A and

[22] R. Suhrmann and G. Barth, Z. Physik **103**, 133 (1936).

[23] R. S. Sennett and G. D. Scott, J. Opt. Soc. Am. **40**, 208 (1950).

increased rapidly at film thicknesses less than 100A. At 44A, a resistivity of 60.8 microhm-cm was obtained which is 25 times the bulk resistivity. Runs 1, 3, 4, and 13 in Table I are not shown in Fig. 3 due to a cracking phenomenon observed in these films as they reached temperatures in the neighborhood of 0°C. The development of the cracks occurred from the surface compression caused by the unequal coefficients of thermal expansion of glass and gold and was accompanied by a sudden and rapid increase in resistance. This phenomenon has been observed previously in connection with x-ray studies[2] in which the x-ray line breadth of the (111) reflections in a gold film suddenly decreased from 1.3 to 1.0 degrees at the same temperature the film cracked and the resistance rose rapidly.

Theoretical

Thomson[24] was the first to propose a formula which expresses the resistivity as a function of film thickness. Other formulas by Planck,[17] Lovell,[13] Appleyard and Lovell,[14] Fuchs,[15] and Weale[16] have also been proposed. None of these can be regarded as completely satisfactory since even the most rigorous ones disagree somewhat with experimental results. In this connection, it must be pointed out that most experimental measurements of electrical resistivity and film thickness are subject to errors arising from the presence of unknown impurities and the difficulties in measuring film thicknesses as small as a few atomic layers.

One of the simplest and less rigorous equations is due to Planck:[17]

$$\rho=\rho_0(1+A/D), \tag{7}$$

where ρ is the resistivity of the film, D the thickness of the film, ρ_0 the resistivity of the bulk metal, and A is a constant given by Weale:[16]

$$A=4l/\pi, \tag{8}$$

where l is the mean free path of the conduction electrons.

The mean free path of the conduction electrons can be calculated from experimental values of bulk resistivity, and was found to be 812A at 0°C for gold.[25]

[24] J. J. Thomson, Proc. Cambridge Phil. Soc. 11, 120 (1901).
[25] N. F. Mott and H. Jones, *The Theory of the Properties of Metals and Alloys* (Oxford University Press, London, 1945), p. 268.

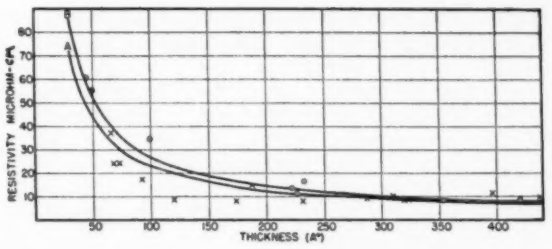


FIG. 3. The final resistivity of gold films as a function of thickness. The circles, represent this research, the x's represent the data of Pogany (see reference 3). Curve A is the Planck-Weale equation with an electron free mean path of 812A and Curve B is the same equation for a free mean path of 970A.

Substituting this value into Eqs. (7) and (8) gives the result:

$$\rho=\rho_0(1+1033/D), \tag{9}$$

which is plotted in Fig. 3 (Curve A) and found to lie slightly lower than our results. Choosing 970A for the mean free path of the conduction electrons gives the equation:

$$\rho=\rho_0(1+1230/D), \tag{10}$$

which is also plotted in Fig. 3 (Curve B). The experimental results are found to be in good agreement with Eq. (10) as shown.

Pogany's results (Fig. 3, crosses) are in considerable disagreement with this research and with the Planck-Weale equation below 250A.

Perrucca[8] has investigated the conductivity of gold films deposited at room temperature and above for thicknesses less than 1000A. He states that the form of the conduction-thickness curve varies with the temperature at which the film is formed, giving a marked point of inflection at 250°C.

Conclusions

It is concluded that resistivities of gold films in thicknesses less than or of the same order of magnitude as the mean free path of the conduction electrons, 970A, must always be somewhat higher than that of the bulk metal and that the presence of occluded nitrogen is a relatively unimportant factor.

High Temperature Structure and Thermal Expansion of Some Metals as Determined by X-Ray Diffraction Data. I. Platinum, Tantalum, Niobium, and Molybdenum*

JAMES W. EDWARDS,† RUDOLPH SPEISER, AND HERRICK L. JOHNSTON
The Cryogenic Laboratory and The Department of Chemistry, The Ohio State University, Columbus, Ohio
(Received August 28, 1950)

The crystal structure and thermal expansion of platinum, tantalum, niobium, and molybdenum have been determined between 1100° and 2500°K. These metals were found to undergo a uniform thermal expansion over the temperature range of this investigation and to undergo no structural change. The permanent elongation of tantalum wires, produced by annealing at temperatures between 2473° and 2773°K, appears to be caused by reorientation of crystal grains in the specimen and to preferred direction of crystal growth during annealing, rather than to a change in crystal structure.

Quadratic equations have been developed for the thermal expansion of platinum, tantalum, niobium, and molybdenum. These equations are represented, respectively, by

$$\Delta a_0/a_0 = 7.543\times10^{-6}(T-291)+2.362\times10^{-9}(T-291)^2,$$
$$\Delta a_0/a_0 = 6.080\times10^{-6}(T-291)+7.50\times10^{-10}(T-291)^2,$$
$$\Delta a_0/a_0 = 7.591\times10^{-6}(T-291)+6.96\times10^{-10}(T-291)^2,$$

and

$$\Delta a^0/a^0 = 0.987\times10^{-3}+2.40\times10^{-6}(T-273)+2.20\times10^{-9}(T-273)^2.$$

Values of the expansion coefficient α were computed for each of the metals by differentiating the above equations. Our values were compared with those obtained from the Grueneisen theory and indicate that the theory provides a reasonably satisfactory means of extrapolating thermal expansion data obtained at low and moderate temperatures.

INTRODUCTION

THE crystal structure and thermal expansion of some metals have become of interest in this Laboratory in connection with a program of high temperature thermodynamics of inorganic substances. Although the crystal structures of platinum, tantalum, niobium, and molybdenum have been determined by earlier investigators,[1] their structures at high temperatures have not been previously investigated.

In the past, the thermal expansions of solids have been measured by several methods. However, the temperatures at which these measurements could be made were limited by the materials of construction of the experimental apparatus. Holborn and Day,[2] and Holborn, Scheel, and Henning[3] made macroscopic measurements of the thermal expansion of platinum up to 1000°C by means of the comparator method. Worthing[4] determined the thermal expansion of tantalum between 300° and 2700°K by the same method. Our research extended the temperature range of platinum, tantalum, and niobium to above 2000°K by use of an x-ray diffraction camera developed at this Laboratory.[5]

TABLE I. Lattice constant and thermal expansion of platinum at various temperatures.

Temp. °K	Lattice constant A	$\Delta a_0/a_0\times10^3$ Expt'l	$\Delta a_0/a_0\times10^3$ Eq. (2)	$\Delta L/L\times10^3$ Holborn and Day	$\Delta L/L\times10^3$ Holborn, Scheel, and Henning
291	3.924				
1073					7.94
1140	3.9564	8.25	8.03	8.62	
1173					9.05
1273					10.19
1335	3.9660	10.7	10.5	10.91	
1565	3.9759	13.2	13.4		
1705	3.9827	15.0	15.4		
1770	3.9877	16.2	16.3		
1770	3.9883	16.4	16.3		
1890	3.9957	18.3	18.1		
2005	4.0023	20.0	19.9		

SAMPLES

Very pure samples of the materials studied were obtained. The American Platinum Works furnished a sample of platinum which they claimed to be better than 99.95 percent pure. The Fansteel Metallurgical Corporation furnished specimens of 99.9 percent pure tantalum and of 99.8 percent pure niobium. An analysis, conducted by the manufacturer, revealed that the principal impurities in the tantalum specimen were carbon and iron, but these were present to less than 0.03 percent. Their analysis of the niobium sample showed that the principal impurity was tantalum which was present to the extent of only 0.06 percent. The Climax Molybdenum Company donated a sample of better than 99.9 percent pure molybdenum which they had obtained from the Fansteel Metallurgical Corporation. The Climax Company reported that our molybdenum sample contained 0.012 percent carbon, and the Fansteel Corporation gave maximum limits of impurities as 0.015 percent carbon, 0.10 percent oxygen, and 0.02 percent nickel.

* This work was supported in part by the ONR under contract with the Ohio State University Research Foundation.

† Gerard Swope Fellow, General Electric Educational Foundation (1948–9).

[1] Ralph W. Wyckoff, *Crystal Structure* (Interscience Publishers, Inc., New York, 1948).

[2] L. Holborn and A. L. Day, Ann. Physik 4, 104 (1901).

[3] Holborn, Scheel, and Henning, Braunschweig, 54 (1919).

[4] A. G. Worthing, Phys. Rev. 28, 198–201 (1926).

[5] Edwards, Speiser, and Johnston, Rev. Sci. Instr. 20, 343–347 (1949).

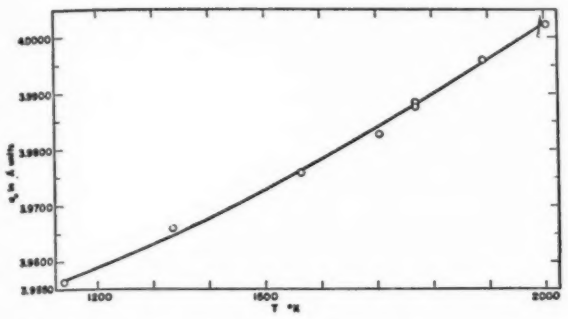


FIG. 1. Lattice constant, a_0, of platinum *vs* temperature.

METHOD

Powder patterns were obtained by means of an x-ray diffraction camera designed for use at high temperatures and developed at this Laboratory.[5] Although the double-walled furnace elements of tantalum, described in an earlier paper,[6] were used in the platinum studies, single-walled elements were used in the other three studies.

Temperatures were measured by means of a Leeds and Northrup optical pyrometer of the disappearing filament type. This pyrometer had been calibrated at 100°C intervals against a standardized lamp obtained from the National Bureau of Standards. The observation window of the diffraction camera was placed between the optical pyrometer and the standard lamp during calibration, thereby obviating the necessity of a separate window correction. The temperatures of the specimen were measured as described earlier;[5] however, the correction scale for the difference in temperatures between the furnace elements and the exposed samples was determined by replacing the sample with a small blackbody in a series of calibration runs. The calibration curve obtained was assumed to apply while the specimen was in position.

Films were measured with a General Electric XRD Film Scale that incorporated a vernier scale which was accurate to 0.005 cm. The films were measured to the centers of the diffraction arcs. Corrections for film shrinkage were made on the assumption of a uniform linear change. Lattice constants, obtained by the Bradley-Jay extrapolation method,[7] were corrected for the index of refraction of the samples. The latter factor was calculated from the equation

$$\Delta d/d = d - d_0/d_0 = S/\sin^2\theta - \zeta = \zeta/\sin^2\theta \qquad (1)$$

proposed by Compton and Allison,[8] where d is the true interplanar spacing, d_0 the spacing calculated by Bragg's law, θ the observed Bragg angle, and ζ the unit decrement of the refractive index or $1-\mu$. This correction amounts to about 0.005 percent. The lattice constants were estimated to be uncertain to about 0.01 percent and the thermal expansion data to about 1 or 2 percent. Lattice constants are given in true angstrom units.[9]

RESULTS

Platinum

Platinum was found to remain face-centered cubic between 1140° and 2005°K. The lattice constants obtained are listed in column 2 of Table I and plotted in Fig. 1, while values for the thermal expansion of platinum, calculated from its lattice constants, are listed in column 3 and plotted in Fig. 2. The latter data can be represented by

$$\Delta a_0/a_0 = 7.543\times10^{-6}(T-291) + 2.362\times10^{-9}(T-291)^2, \qquad (2)$$

where a_0 is the lattice constant at 291°K and Δa_0 is the change in lattice constant between 291°K and some higher temperature T. This equation is applicable over the temperature range 1140° to 2005°K, with a mean deviation of $\pm0.19\times10^{-3}$. These values agree well with those of Holborn and Day,[2] and of Holborn, Scheel, and Henning,[3] as is shown in Fig. 2.

In our studies with platinum, the measured lattice constants were somewhat erratic due to recrystallization occurring at elevated temperatures. As a result, measurements of positions of diffraction arcs were subject to more error than would be expected with a material such as tantalum.

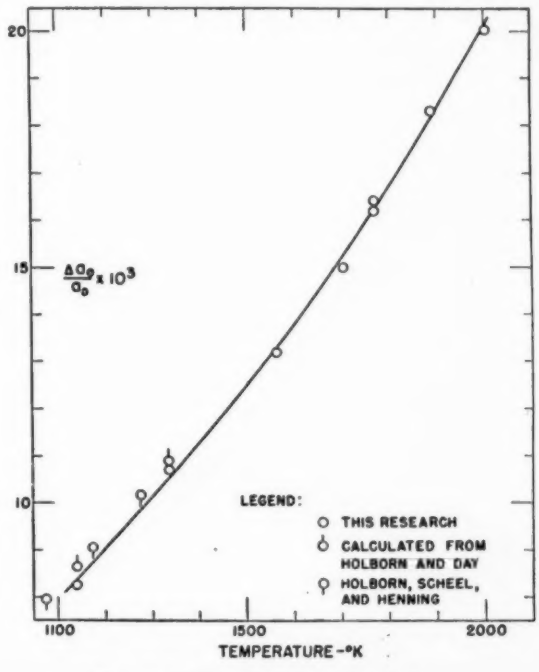


FIG. 2. Thermal expansion of platinum.

[6] Edwards, Speiser, and Johnston, Rev. Sci. Instr. **20**, 367 (1949).
[7] A. J. Bradley and A. H. Jay, Proc. Phys. Soc. (London) **44**, 563 (1932).
[8] A. H. Compton and S. K. Allison, *X-Rays in Theory and Experiment* (D. Van Nostrand and Company, Inc., New York, 1935), p. 674.
[9] Elizabeth Armstrong Wood. Phys. Rev. **72** 436 (1947).

TABLE II. Lattice constants and thermal expansion of tantalum at various temperatures.

Temp. °K	Lattice constant A	$\Delta a_0/a_0 \times 10^3$ Expt'l	Eq. (3)	$\Delta L/L \times 10^3$ Worthing
291	3.3029	0	0	—
1135	3.3218	5.76	5.67	5.87
1281	3.3254	6.86	6.75	6.97
1454	3.3295	8.10	8.09	8.31
1641	3.3342	9.53	9.58	9.79
1826	3.3397	11.2	11.1	11.3
2030	3.3453	12.9	12.8	13.0
2091	3.3490	14.0	13.4	13.5
2299	3.3518	14.8	15.2	15.3
2495	3.3581	16.7	17.0	17.0

Tantalum

Tantalum was found to be body-centered cubic from 1135° to 2495°K, just as at room temperature. The lattice constants obtained are listed in column 2 of Table II, while values for the thermal expansion of tantalum, calculated from its lattice constants, are listed in column 3 and are plotted in Fig. 3. The latter data are represented by

$$\Delta a_0/a_0 = 6.080\times10^{-6}(T-291) + 7.50\times10^{-10}(T-291)^2, \quad (3)$$

where the symbols have the same significance as in Eq. (2). This equation is applicable from 1135° to 2495°K, with a mean deviation of $\pm 0.20\times10^{-3}$.

Worthing obtained the equation

$$\Delta L/L = 6.60\times10^{-6}(T-300)+5.2\times10^{-10}(T-300)^2 \quad (4)$$

from his thermal expansion measurements over the temperature range 300° to 2700°K. In his equation, L is the length of a well-seasoned tantalum wire at 300°K, and ΔL is the change in length between 300°K and some higher temperature T. Values of thermal expansion, calculated from Eq. (4), are listed in column 5 of Table II.

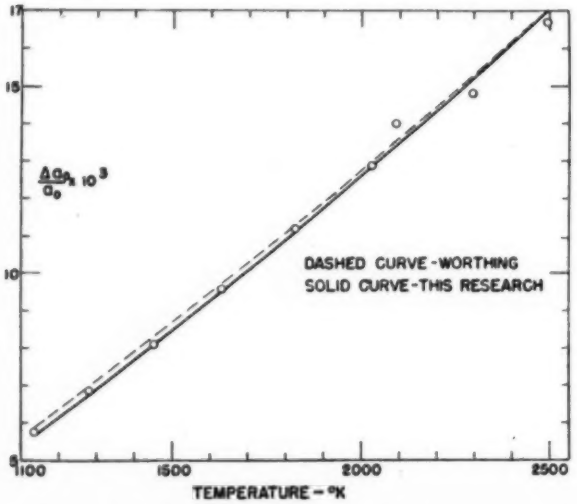


FIG. 3. Thermal expansion of tantalum.

Webber,[10] whose tantalum sample was of the same source and purity as ours, reported a permanent over-all elongation of 2 percent after heating his wire for two hours at temperatures between 2473° and 2773°K and cooling to room temperature before making measurements. This result compares with our x-ray results of an over-all thermal expansion of the lattice of only 1.67 percent between room temperature and 2495°K, with no permanent lattice expansion at room temperature after recooling. Worthing, who used an annealed sample of tantalum wire, observed no permanent elongation upon cooling even though he also reported a reversible over-all elongation of 1.70 percent between room temperature and 2495°K.

These apparently discrepant results can be explained if it is assumed that the cold-drawn wires used by Webber and by us were originally in a strained condition and that the strain was relieved by some rearrangement of crystallites which would increase the over-all length of the wire and reduce its diameter during high temperature recrystallization. This assumption would explain why the permanent elongation reported for Webber's wire could not occur in Worthing's annealed wire which was free of strain. It would not be observed in our measurements of lattice constants since our measurements were on the individual crystallites and not on the over-all wire.

TABLE III. Lattice constants and thermal expansion of niobium at various temperatures.

Temp. °K	Lattice constant A	$\Delta a_0/a_0 \times 10^3$ Expt	Eq. (5)	$a_0(\text{II}) - a_0(\text{I})$ A $\times 10^4$
291	3.3004	—	—	—
1115	3.3234	6.80	6.73	16
*1225	3.3266	7.77	7.70	34
*1270	3.3298	8.77	8.10	37
*1355	3.3312	9.17	8.86	31
*1360	3.3287	8.41	8.91	9
1360	3.3295	8.65	8.91	10
*1465	3.3367	10.9	9.87	31
1495	3.3342	10.1	10.2	10
1655	3.3400	11.8	11.7	8
1855	3.3457	13.6	13.6	8
1870	3.3463	13.8	13.7	17
*1965	3.3500	14.9	14.7	39
2116	3.3543	16.2	16.2	—
2309	3.3609	18.2	18.2	—
2470	3.3663	19.8	19.9	—

Niobium

The structure of niobium was found to be body-centered cubic from 1115° to 2474°K, just as at room temperature. Its lattice constants and thermal expansion data are given in Table III. The thermal expansion of niobium can be represented by

$$\Delta a_0/a_0 = 7.591\times10^{-6}(T-291)+6.96\times10^{-10}(T-291)^2, \quad (5)$$

where the symbols are defined as in Eq. (2). This equa-

[10] R. T. Webber, Phys. Rev. **72**, 1241 (1947).

tion is applicable from 1115° to 2474°K, with a mean deviation of $\pm 0.09\times10^{-3}$. Certain data, marked * in Table III and shown in triangles in Fig. 4, were not used in deriving Eq. (5) because an excessively large residual lattice expansion existed at room temperature after these high temperature runs. This is exhibited in column 5 in which a_0(II) is the room temperature lattice constant of the specimen after heating in the camera and a_0(I) its room temperature value before heating. This residual expansion, attributed to the solution of gaseous impurities in solid niobium, was decreased by preheating in vacuum, in the high temperature camera prior to exposure. The thermal expansion of niobium is presented graphically in Fig. 4, wherein the circles and triangles represent experimental points, and the curve Eq. (5).

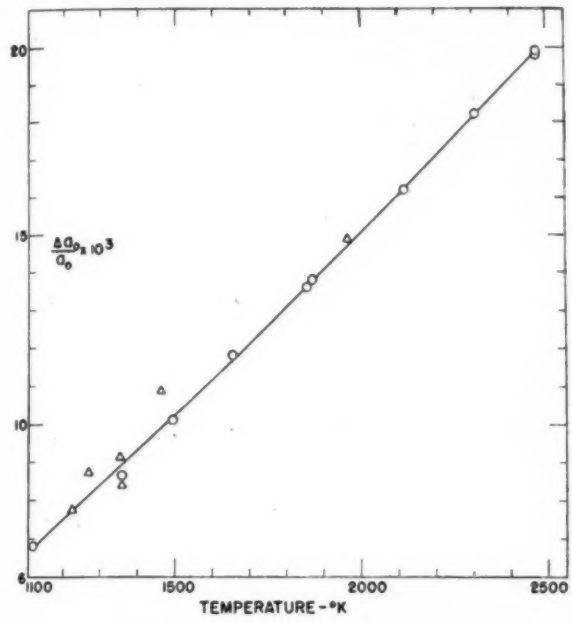


FIG. 4. Thermal expansion of niobium.

Molybdenum

These experiments showed that molybdenum remained body-centered cubic from 1126° to 2073°K, just as at room temperature. The lattice constants obtained are listed in column 2 of Table IV along with those calculated from the data of Worthing[4] and of Demarquay.[11] The values for the thermal expansion of molybdenum, calculated from its lattice constants, are listed in column 3 and are plotted in Fig. 5 along with the data of Worthing[4] and Demarquay.[11] Our thermal expansion data can be represented by

$$\Delta a_0/a_0=0.987\times10^{-3}+2.40+10^{-6}(T-273)+2.20\times10^{-9}(T-273)^2, \quad (6)$$

where a_0 is the lattice constant at 273°, and Δa_0 is the change in lattice constant between 273°K and some higher temperature T. Since Mo_2C separated to form two solid phases upon annealing between 1700° and 1800°K, this equation is applicable only from 1126° to 1819°K. Above 1819°K the diffraction pattern for molybdenum was blurred, and therefore the data at higher temperatures are not considered reliable. Between 1126° and 1819°K, Eq. (6) is applicable within a mean deviation of $\pm 0.08\times10^{-3}$.

[11] Jean Demarquay, Compt. rend. **220**, 81–83 (1945).

TABLE IV. Lattice constants and thermal expansion of molybdenum at various temperatures.

Temp. °K	Lattice constants A	$\Delta a_0/a_0\times10^3$ Expt.	Eq. (6)	$\Delta L/L\times10^3$ Worthing	Demarquay
291	3.1474	—	—	—	—
1126	3.1621	4.65	4.64	4.85	4.62
1162	3.1623	4.71	4.86	5.09	4.84
1327	3.1668	6.11	5.96	6.24	5.90
1510	3.1706	7.35	7.32	7.59	7.13
1639	3.1738	8.34	8.37	8.58	8.18
1819	3.1782	9.74	9.96	10.02	9.59
1968	3.1832	11.3	11.38	11.26	10.96
2073	3.1869	12.5	12.44	12.17	12.06

RELIABILITY OF THE GRUENEISEN EQUATION

Grueneisen[12] developed the approximate equation

$$\alpha=\frac{C_v}{3Q_0[1-(m+n+3/6)(E/Q)]^2} \quad (7)$$

for the linear thermal expansion coefficient of solids, wherein C_v is the atomic heat capacity at constant volume, E equals $\int_0^T C_v dT$, Q_0 is a constant obtained from a zero-order approximation of the relation of

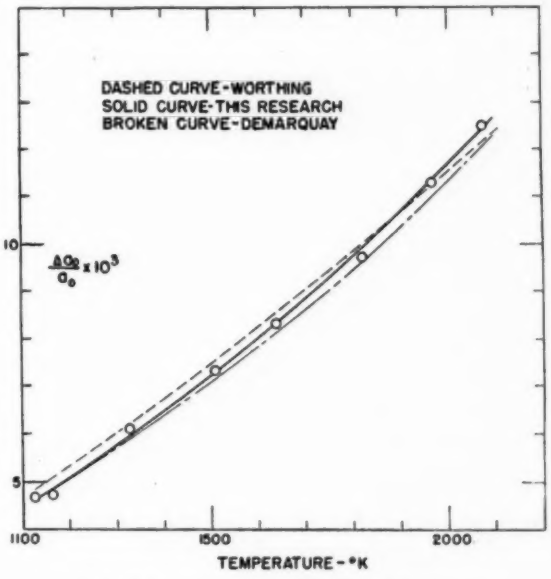


FIG. 5. Thermal expansion of molybdenum.

[12] E. Grueneisen, "Thermische Eigenschaften der Stoffe." *Handbuch der Physik* (Verlag Julius Springer, Berlin, 1926), Vol. X, pp. 1–59.

TABLE V. Thermal expansion coefficients of platinum, tantalum, niobium, and molybdenum.

Temp. °K	$\alpha \times 10^6$ Pt Expt.[a]	Pt Gruen.[b]	Ta Expt.[a]	Ta Gruen.[b]	Cb Expt.[a]	Cb Gruen.[b]	Mo Expt.[a]	Mo Gruen.[b]	Mo Worth.[a]
1000	—	—	—	—	—	—	5.61	7.70	6.47
1100	11.4	13.0	7.29	7.63	8.72	8.72	6.05	—	—
1200	11.8	13.6	—	—	—	—	6.49	8.35	6.89
1300	12.3	14.1	7.59	7.79	9.00	8.94	—	—	—
1400	12.8	14.7	—	—	—	—	7.37	9.16	7.31
1500	13.3	15.6	7.89	7.97	9.27	9.20	—	—	—
1600	13.7	16.0	—	—	—	—	8.26	10.07	7.73
1700	14.2	16.7	8.19	8.07	9.55	9.54	—	—	—
1800	14.7	17.4	—	—	—	—	9.14	11.10	8.15
1900	15.1	18.2	8.49	8.36	9.83	9.83	—	—	—
2000	15.6	19.0	—	—	—	—	10.02	12.47	8.57
2100	—	—	8.79	8.60	10.11	10.06	—	—	—
2200	—	—	—	—	—	—	10.90	—	9.99
2300	—	—	9.09	8.86	10.39	10.22	—	—	—
2500	—	—	9.39	9.12	10.67	10.36	—	—	—

[a] This research.
[b] Calculated using Grueneisen's theory.
[c] Calculated from Worthing's equation.

atomic volume to total energy of a crystal, and m and n are exponents of an assumed potential function. The approximations made in developing Eq. (7) limit the agreement obtainable between the results calculated from the equation and actual experimental data. However, Grueneisen[12] and Nix and MacNair[13] found that empirical values could be obtained for the constants Q_0 and $(m+n+3)/6$ by fitting the equation to their experimental data. They found that the equation then closely represented the change of the thermal expansion coefficient at low temperatures. They obtained good agreement between the calculated and the experimental values over wide ranges of temperature, particularly in the low temperature region where both the thermal expansion coefficients and the heat capacities change rapidly with temperature.

The thermal expansion coefficients obtained from Eqs. (2), (3), (5), and (6) are compared in Table V with values obtained by Grueneisen's theory. The values of Q_0 and $(m+n+3)/6$, 231×10^3 cal and 4.8, respectively, for the platinum calculations, are taken from Grueneisen's paper,[12] while those for tantalum, 292.4×10^3 cal and 0.2924, and for molybdenum, 365×10^3 cal and 6.83, are those obtained empirically by Nix and MacNair[13] by fitting the Grueneisen equation to their experimental data between 195° and 673°K. The constants used in the niobium calculations were determined by fitting our experimental data to the Grueneisen equation. In all four calculations, the C_v values used in Eq. (7) were obtained from smoothed values for C_p given by Simon and Zeidler,[14] by Cooper and Langstroth,[15] by Bronson and Chisholm,[16] and by Jaeger and Veenstra.[17] The latter data were extrapolated to the temperature of our investigation by the approximate relation

$$aC_p^2T=C_p-C_v, \tag{8}$$

where a is an empirical constant. The values of E, used in Eq. (7), were obtained by graphical integration of $C_v dT$. These calculations show that the Grueneisen theory provides a satisfactory means of obtaining thermal expansion data for tantalum and for niobium at high temperatures by extrapolation of low temperature data. In addition, the agreement between our calculated thermal expansion coefficient for platinum and molybdenum and those calculated from Grueneisen's theory was reasonable in view of the difficulties encountered with platinum recrystallizing and with Mo_2C forming a second phase.

[13] F. C. Nix and D. MacNair, Phys. Rev. **61**, 74–78 (1942).

[14] F. Simon and W. Zeidler, Z. physik. Chem. **123**, 384–404 (1926).
[15] D. Cooper and G. O. Langstroth, Phys. Rev. **33**, 243–248 (1929).
[16] H. L. Bronson and H. M. Chisholm, J. Inst. Metals **47**, 261 (1931).
[17] F. M. Jaeger and W. A. Veenstra, Rec. trav. chim. **53**, 677–687 (1934).

Midwest Research Institute

An annual report of operations which incorporates the most progressive advances made since its inception has been released by Midwest Research Institute, Kansas City.

Increased number of projects, expanded facilities, augmented staff, wider degree of future research commitments, and a continuous improvement of scientific aptitudes sparked this sixth year of Institute activity. Founded originally to serve a six-state area, the Institute now services agricultural and industrial interests from coast to coast and in Latin America.

The 130 staff members of the three Institute divisions, Chemistry, Physics, and Engineering, utilize over one-half million dollars in technical equipment to administer the diverse needs of small industrial clients with no research facilities of their own, as well as larger organizations who are desirous of a "fresh approach" to a specific problem.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 4 APRIL, 1951

The Oscillations of Magnetic Suspensions

KNOX MILLSAPS AND J. C. MCPHERSON
Department of Physics, Alabama Polytechnic Institute, Auburn, Alabama
(Received August 28, 1950)

The oscillations of a magnetic suspension between two equal poles are formulated in terms of a nonlinear ordinary differential equation of the second order that may be integrated in terms of elliptic integrals of the second kind. The case of unequal poles is treated by an analogous procedure involving a Legendre reduction. The further refinement of viscous damping is treated by numerical integration, and the relaxation times are found by exponential extrapolation.

INTRODUCTION

MAGNETIC suspensions are often used when it is desirable to reduce the frictional level in mechanisms to an absolute minimum. The most important instances, at present, are in electric meters, analytical balances, and gyroscopes that are used in obtaining a reference to inertial space in ultra-long range guidance systems.

Although a study of the vibratory characteristics of such systems proves to be quite elementary—in fact, it seems to provide a very elementary physical example whose mathematical analysis uses elliptic integrals of the second kind,—it is not found in the standard reference books on nonlinear oscillations[1] or applications of elliptic functions;[2] hence, it appears worthwhile to make a theoretical investigation of the oscillations of magnetic suspensions.

In this paper, an analysis is made of the nonlinear oscillations of a unidimensional magnetic suspension. After deriving the general equation of motion, it is convenient to divide the analysis into three parts: the motion in vacuo between equal poles, the motion in vacuo between unequal poles, and the motion between equal poles with heavy and light viscous damping. It is possible to treat the first two cases by exact methods with the integrals given explicitly in finite form. In the remaining case, it does not seem possible to obtain an explicit solution in finite form, and, consequently, a numerical analysis is made.

I. THE EQUATION OF MOTION

In order to derive the differential equation of motion for the oscillation of a unidimensional magnetic suspension, it is convenient to geometrize the analysis. A magnetic pole P_2 (unit poles) is suspended between two fixed magnetic poles P_1 and P_3 (unit poles) according to Fig. 1. In Fig. 1, 2a is the distance between P_1 and P_3 (cm), e is the distance from P_1 to the equilibrium position (cm), ξ is the displacement from the equilibrium position, (cm), and ξ_0 is the maximum (usually the initial), displacement (cm).

[1] A. A. Andronow and C. E. Chaikin, *Theory of Oscillations* (Princeton University Press, Princeton, New Jersey, 1949).
[2] W. Magnus and F. Oberhettinger, *Andwendung der elliptischen Funktionen in Physik und Technik* (Springer-Verlag., Berlin, Germany, 1949).

The ordinary second-order differential equation describing the oscillation of such a suspension is

$$m\frac{d^2\xi}{dt^2}=\frac{P_1P_2}{(e+\xi)^2}-\frac{P_2P_3}{(2a-e-\xi)^2}-\Delta\frac{d\xi}{dt}\left|\frac{d\xi}{dt}\right|, \qquad (1.1)$$

where m is the mass of P_2 (g), Δ is the coefficient of drag in an incompressible viscous medium (g cm^{-1}), and t is the time (sec).

The differential equation may be put into a dimensionless form by the following transformations:

$\alpha=\xi\xi_0^{-1}$, giving a dimensionless amplitude, (1)

$\delta=2\Delta m^{-1}\xi_0$, giving a dimensionless damping, (2)

$\epsilon=e\xi_0^{-1}$, giving a dimensionless equilibrium position, (3)

$\rho=P_3P_1^{-1}$, giving a ratio of pole strengths, (4)

$\sigma=a\xi_0^{-1}$, giving a dimensionless spacing, (5)

and

$\tau=2t[aP_1P_2]^{\frac{1}{2}}m^{-\frac{1}{2}}\xi_0^{-2}$, giving a dimensionless time. (6)

The transformed equation is

$$\frac{d^2\alpha}{d\tau^2}=\frac{1}{4\sigma}\left[\frac{1}{(\epsilon+\alpha)^2}-\frac{\rho}{(2\sigma-\epsilon-\alpha)^2}\right]-\frac{\delta}{2}\frac{d\alpha}{d\tau}\left|\frac{d\alpha}{d\tau}\right|. \qquad (1.2)$$

Obviously, ϵ, σ, and ρ are not independent. Between them there exists the functional relation,

$$\epsilon=\frac{2\sigma}{1-\rho}[1-\rho^{\frac{1}{2}}]. \qquad (1.3)$$

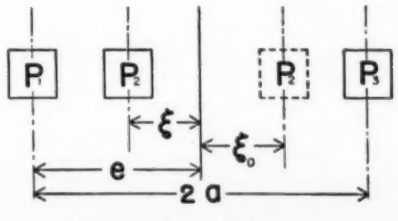


FIG. 1. Schematic diagram.

II. OSCILLATIONS IN VACUO BETWEEN EQUAL POLES

The ideal case of oscillations between equal poles in an inviscid medium may be treated by putting $\delta=0$, $\epsilon=\sigma$, and $\rho=1$. The equation of motion for this simplified case is

$$\frac{d^2\alpha}{d\tau^2}=\frac{1}{4\sigma}\left[\frac{1}{(\sigma+\alpha)^2}-\frac{1}{(\sigma-\alpha)^2}\right]. \tag{2.1}$$

The independent variable τ does not occur explicitly in this equation, so, in accordance with a familiar procedure, the equation is reduced to

$$\varphi\frac{d\varphi}{d\alpha}=\frac{1}{4\sigma}\left[\frac{1}{(\sigma+\alpha)^2}-\frac{1}{(\sigma-\alpha)^2}\right] \tag{2.2}$$

by the transformation

$$d\alpha/d\tau=\varphi. \tag{2.3}$$

Integrating and using the initial condition,

$$\varphi=0 \quad \text{at} \quad \xi=\xi_0, \tag{2.4}$$

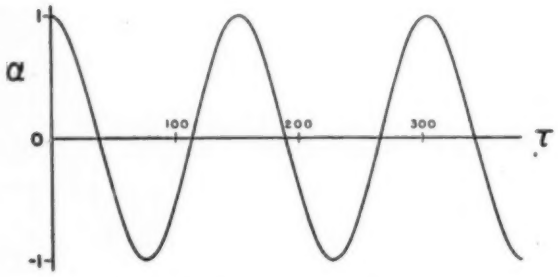


FIG. 2. Oscillations between equal poles with $\delta=0$, $\rho=1$, and $\sigma=5$.

the equation may be put in the following form:

$$\tau=\sigma(\sigma^2-1)^{\frac{1}{2}}\int_a^1\frac{(1-\sigma^{-2}\alpha^2)^{\frac{1}{2}}}{(1-\alpha^2)^{\frac{1}{2}}}d\alpha, \tag{2.5}$$

where on taking the square root the negative sign has been retained in order to obtain a physically significant solution.

The indicated quadrature may be performed in terms of elliptic integrals of the second kind;[3] the solution is

$$\tau=\sigma(\sigma^2-1)^{\frac{1}{2}}[E(\sigma^{-1},\pi/2)-E(\sigma^{-1},\text{arc sin}\alpha)]. \tag{2.6}$$

Figure 2 shows the amplitude α as a function of τ with $\sigma=5$.

III. OSCILLATIONS BETWEEN UNEQUAL POLES

The characteristics of the oscillations for an asymmetric suspension are most easily considered by assuming $\delta=0$. For this case, the general equation becomes

$$\frac{d^2\alpha}{d\tau^2}=\frac{1}{4\sigma}\left[\frac{1}{(\epsilon+\alpha)^2}-\frac{\rho}{(2\sigma-\epsilon-\alpha)^2}\right]. \tag{3.1}$$

Once again the independent variable does not occur explicitly; hence, the equation may be reduced to one of the first order. Fortunately, the variables in the first-order equation are easily separated. The physically significant solution in quadrature form is

$$\tau=[2\sigma(A+1)(B-1)]^{\frac{1}{2}}\times\int_a^1\frac{\gamma_0\alpha^2+\gamma_1\alpha+1}{[\beta_0\alpha^4+4\beta_1\alpha^3+6\beta_2\alpha^2+4\beta_3\alpha+\beta_4]^{\frac{1}{2}}}d\alpha, \tag{3.2}$$

where

$$\gamma_0=-(1/AB),$$
$$\gamma_1=(B-A)/AB,$$
$$\beta_0=C+D,$$
$$4\beta_1=2AD-2BC+AC-BD+C+D,$$
$$6\beta_2=A^2D+B^2C-2ABD-2ABC-2AD+2BC-AC+BD,$$
$$4\beta_3=AB^2C-A^2BD-A^2D-B^2C+2ABD+2ABC,$$

and

$$\beta_4=A^2BD-AB^2C.$$

The secondary series of constants, A, B, C, and D, are related to the initial parameters in the following way:

$$A=\epsilon,$$
$$B=2\sigma-\epsilon,$$
$$C=2\sigma-\epsilon-1,$$

and

$$D=\rho(\epsilon+1).$$

If the roots of the denominator of the integrand in Eq. (3.2) are denoted by $a>b>c>d$, Eq. (3.2) may be integrated in terms of elliptic functions;[4] the exact solution is

$$\tau=\{AB[2\sigma(A+1)(B-1)]^{\frac{1}{2}}\}\left\{\gamma_0\beta_0^{-1}[-\beta_1\mu-\zeta(\mu+\nu)+\zeta(\mu)]+[\gamma_1-2\beta_0^{-1}\beta_1\gamma_0]\times[(\beta_0^{-1}\beta_1+\beta_0^{-\frac{1}{2}}\zeta(\nu))]+\beta_0^{-\frac{1}{2}}\ln\frac{\sigma(\mu+\nu)}{\sigma(\mu)}+\mu\right\}, \tag{3.3}$$

where σ and ζ are the familiar Weierstrass functions.[5] In addition,[6]

$$\mu=2[(a-c)(b-d)]^{-\frac{1}{2}}F(k,\alpha), \tag{3.4}$$

where $F(k,\alpha)$ is the Legendre elliptic integral of the

[3] E. Jahnke and F. Emde, *Funktionentafeln* (B. G. Teubner, Leipzig, Germany, 1933), p. 54.

[4] J. Tannery and J. Molk, *Théorie des fonctions elliptiques* (Gauthier-Villars, Paris, France, 1893–1902), tome IV, p. 156, formulas CXLIV (1), (2), and (3).

[5] J. Tannery and J. Molk, reference 4, tome II, p. 234.

[6] J. Tannery and J. Molk, reference 4, tome II, p. 108.

first kind, and k is the elliptic modulus given by $k^2=(b-c)(a-d)(a-c)^{-1}(b-d)^{-1}$ and, clearly, ν is the elliptic periodicity.[7]

IV. OSCILLATIONS IN VISCOUS MEDIA

In order to avoid undue complications, it is convenient to consider only the case of viscous damping between equal poles; therefore, it is assumed that $\epsilon=\sigma$ and $\rho=1$. The general equation then becomes

$$\frac{d^2\alpha}{d\tau^2}=\frac{1}{4\sigma}\left[\frac{1}{(\sigma+\alpha)^2}-\frac{1}{(\sigma-\alpha)^2}\right]-\frac{\delta}{2}\frac{d\alpha}{d\tau}\left|\frac{d\alpha}{d\tau}\right|. \tag{4.1}$$

Now it is advantageous to consider two cases: (A), $d\alpha/d\tau\leqslant 0$ and (B), $d\alpha/d\tau\geqslant 0$.

Case (A):

When $d\alpha/d\tau\leqslant 0$, Eq. (4.1) becomes

$$\frac{d^2\alpha}{d\tau^2}-\frac{\delta}{2}\left(\frac{d\alpha}{d\tau}\right)^2=\frac{1}{4\sigma}\left[\frac{1}{(\sigma+\alpha)^2}-\frac{1}{(\sigma-\alpha)^2}\right]. \tag{4.2}$$

If one puts

$$(d\alpha/d\tau)^2=Z, \tag{4.3}$$

Eq. (4.2) reduces to

$$\frac{dZ}{d\alpha}-\delta Z=\frac{1}{2\sigma}\left[\frac{1}{(\sigma+\alpha)^2}-\frac{1}{(\sigma-\alpha)^2}\right]. \tag{4.4}$$

Equation (4.4) is an ordinary differential equation of the first order with the well-known solution,

$$\frac{d\alpha}{d\tau}=[2\sigma e^{-\delta\alpha}]^{-\frac{1}{2}}\left[-\frac{e^{-\delta\alpha}}{\sigma+\alpha}-\frac{e^{-\delta\alpha}}{\sigma-\alpha}-\delta\int\frac{e^{-\delta\alpha}}{\sigma+\alpha}d\alpha-\delta\int\frac{e^{-\delta\alpha}}{\sigma-\alpha}d\alpha\right]^{\frac{1}{2}}. \tag{4.5}$$

With sufficient accuracy for small damping, the exponential integral may be approximated by the equation,[8]

$$\int\frac{e^{-\delta\alpha}}{\sigma+\alpha}d\alpha=e^{\delta\sigma}\left[\ln\delta(\sigma+\alpha)-\delta(\sigma+\alpha)+\frac{\delta^2(\sigma+\alpha)^2}{4}\right]. \tag{4.6}$$

With a little algebraic manipulation, the substitution of Eqs. (4.3) and (4.6) into Eq. (4.5) gives

$$\tau=\int_{+C_i}^{\alpha}[f_1(C_i)-f_1(\alpha)]^{-\frac{1}{2}}d\alpha, \tag{4.7}$$

where

$$f_1(\alpha)=[2\sigma e^{-\delta\alpha}]^{-1}\left\{-\frac{e^{-\delta\alpha}}{\sigma+\alpha}-\frac{e^{-\delta\alpha}}{\sigma-\alpha}+\delta e^{\delta\sigma}[\ln\delta(\sigma+\alpha)-\delta(\sigma+\alpha)+\delta^2(\sigma+\alpha)^2/4]-\delta e^{-\delta\sigma}[\ln\delta(\sigma-\alpha)+\delta(\sigma-\alpha)+\delta^2(\sigma+\alpha)^2/4]\right\}. \tag{4.8}$$

The $f_1(C_i)$ appearing in Eq. (4.7) is introduced by the initial condition,

$$d\alpha/d\tau=0 \quad \text{at} \quad \alpha=C_i. \tag{4.9}$$

It is clear that $f_1(C_i)-f_1(\alpha)$ vanishes for two values, $\alpha=C_i$ and $\alpha=-C_{i+1}$ and that $C_i\geqslant C_{i+1}$.

The integration of Eq. (4.7) in finite terms is not immediately evident. Thus, it seems advisable to use numerical integration; however, it is noted that the integrand has singularities at $\alpha=C_i$ and $\alpha=C_{i+1}$. One way that has been suggested to eliminate the difficulty at the singularities is to use the quadratic approximation,

$$f_1(C_i)-f_1(\alpha)=K_1(C_i-\alpha)(C_{i+1}+\alpha), \tag{4.10}$$

where

$$K_1=\frac{\max[f_1(C_i)-f_1(\alpha)]}{C_iC_{i+1}+(C_i-C_{i+1})^2/4}. \tag{4.11}$$

If the transformation,

$$\varphi=(C_i+C_{i+1})^{-1}\ln[(C_{i+1}+\alpha)(C_i-\alpha)^{-1}], \tag{4.12}$$

is used, Eq. (4.7) becomes

$$\tau=K_1^{-\frac{1}{2}}\int\left[f_1(C_i)-f_1\left(\frac{C_i\exp[\varphi(C_i+C_{i+1})]-C_{i+1}}{1+\exp[\varphi(C_i+C_{i+1})]}\right)\right]^{-\frac{1}{2}}d\alpha \tag{4.13}$$

with the limits of integration $+\infty$ to

$$\frac{C_i\exp[\varphi(C_i+C_{i+1})]-C_{i+1}}{1+\exp[\varphi(C_i+C_{i+1})]}.$$

Case (B):

In a manner analogous to that employed in Case (A), one easily obtains

$$\tau=\int_{C_{i+1}}^{\alpha}[f_2(C_{i+1})-f_2(\alpha)]^{-\frac{1}{2}}d\alpha, \tag{4.14}$$

where

$$f_2(\alpha)=[2\sigma e^{+\delta\alpha}]^{-1}\left\{-\frac{e^{\delta\alpha}}{\sigma+\alpha}-\frac{e^{\delta\alpha}}{\sigma-\alpha}+\delta e^{-\delta\sigma}[\ln\delta(\sigma+\alpha)+\delta(\sigma+\alpha)+\delta^2(\sigma+\alpha)^2/4]-\delta e^{-\delta\sigma}[\ln\delta(\sigma-\alpha)-\delta(\sigma-\alpha)+\delta^2(\sigma-\alpha)^2/4)]\right\}. \tag{4.15}$$

If the approximation,

$$f_2(C_{i+1})-f_2(\alpha)=K_2(C_{i+1}+\alpha)(C_{i+2}-\alpha), \tag{4.16}$$

where

$$K_2=\frac{\max[f_2(C_{i+1})-f_2(\alpha)]}{C_{i+1}C_{i+2}+(C_{i+1}-C_{i+2})^2/4}, \tag{4.17}$$

and a similar transformation,

$$\varphi=(C_{i+1}+C_{i+2})^{-1}\ln[(C_{i+1}+\alpha)(C_{i+2}-\alpha)^{-1}], \tag{4.18}$$

[7] J. Tannery and J. Molk, reference 4, tome II, p. 235.
[8] G. P. Bois, *Tafeln Unbestimmter Integrale* (B. G. Teubner, Leipzig, Germany, 1906), p. 145.

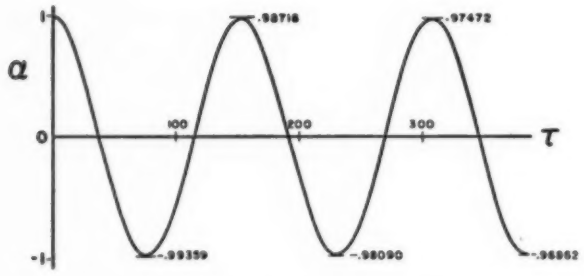


FIG. 3. Oscillations between equal poles with $\delta=.01$, $\rho=1$, and $\sigma=5$.

are used, Eq. (4.14) becomes

$$\tau=K_2^{-\frac{1}{2}}\int\Big[f_2(C_{i+1})-f_2\Big(\frac{C_{i+2}\exp[\varphi(C_{i+1}+C_{i+2})]-C_{i+1}}{1+\exp[\varphi(C_{i+1}+C_{i+2})]}\Big)\Big]^{-\frac{1}{2}}d\alpha \quad (4.19)$$

with the limits of integration, $-\infty$ to

$$\frac{C_{i+2}\exp[\varphi(C_{i+1}+C_{i+2})]-C_{i+1}}{1+\exp[\varphi(C_{i+1}+C_{i+2})]}.$$

The characteristics of the damped motion may be obtained by numerically integrating Eqs. (4.14) and (4.10) using Weddle's rule;[9] the results are given in Fig. 3. It is noted that a very good approximation to the curve for damped motion is made by fitting the values for undamped motion between the appropriate singularities.

In general, the calculation of the decay curves for damped motion required the repeated solution of transcendental iterative equations and is far too lengthy a process to use in a mathematical description of the decay of the initial vibration; however, it is instructive to consider the first few roots. In Fig. 4 decay curves

[9] J. B. Scarborough, *Numerical Mathematical Analysis* (Oxford University Press, London, 1930), p. 120.

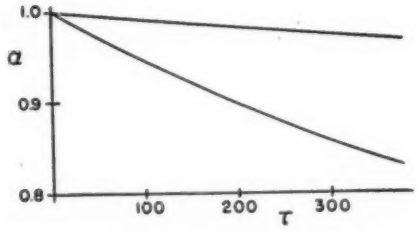


FIG. 4. Decay curves for short intervals of τ with different dampings. For the upper curve $\delta=0.01$, $\rho=1$, and $\sigma=5$; for the lower curve $\delta=0.1$, $\rho=1$, and $\sigma=5$.

are plotted for short time intervals to show the effects of different viscosities.

The relaxation time, $\lambda=\tau(\alpha_0\varepsilon^{-1}, \delta)$, for these oscillations is obtained by exponential extrapolation. The decay curves are assumed to be

$$\exp(\beta_1\tau+\beta_2\tau^2+\beta_3\tau^3), \quad (4.20)$$

where

$$n\beta_i=\textstyle\sum_j\beta_{ij} \quad (i=1, 2, 3; j=1, \cdots, n),$$
$$\beta_{1j}=d^{-1}\tau_1^5[(j+1)^2(j+2)^2\ln C_j - 2j^2(j+2)^2\ln C_{j+1}+j^2(j+1)^2\ln C_{j+2}],$$
$$\beta_{2j}=d^{-1}\tau_1^4[-(j+1)(j+2)(j+3)\ln C_j + 4j(j+1)(j+2)\ln C_{j+1}-j(j+1)(2j+1)\ln C_{j+2}],$$
$$\beta_{3j}=d^{-1}\tau_1^3[(j+1)(j+2)\ln C_j-2j(j+2)\ln C_{j+1} + j(j+1)\ln C_{j+2}],$$
$$d=2\tau_1^6 j(j+1)(j+2),$$

and τ_1 is one-half of the period. Taking $n=6$ for the two previously considered cases, the relaxation times are found to be $\lambda_{0.1}=2361$ and $\lambda_{0.01}=5749$.

ACKNOWLEDGMENTS

The authors wish to thank Dr. Karl Pohlhausen of the Office of Air Research for suggesting the problem and the Graduate School of the Alabama Polytechnic Institute for making a grant in aid of the research. They also wish to acknowledge the aid given by Messrs. J. B. Dozier, W. B. Ard, and F. G. Manley in performing the laborious computations for the case of viscous damping.

Analysis of Synchronous Conditions in the Cylindrical Magnetron Space Charge*

H. W. WELCH, JR., AND W. G. DOW
Engineering Research Institute, University of Michigan, Ann Arbor, Michigan
(Received August 31, 1950)

In the multianode cylindrical magnetron there exist favored phase velocities of the electromagnetic wave around the interaction space between anode and cathode. These velocities are characteristic of the resonant system attached to the anode segments. In the oscillating magnetron the electronic space charge within the interaction space is presumed to maintain sychronism with one of these velocities. Certain of the conditions of synchronism which can be discussed analytically are treated in this paper. The results, although based on restrictive assumptions, can be used in the interpretation of magnetron operation and in predicting regions of efficient behavior.

INTRODUCTION

THE analysis of space charge distribution in the static magnetron diode has been discussed in a number of papers[1] and has admittedly only been treated under very special assumptions. In spite of this the results can be very useful in many practical applications if properly handled. The resulting picture is essentially that of a swarm of electrons moving around the cathode having a radial extent determined by the applied dc anode voltage and axial magnetic field.

This paper discusses a similar solution for a magnetron in which an rf voltage is assumed present between anode segments uniformly spaced about the magnetron anode periphery. A voltage of this type varying in time has been considered by Slater and others as the resultant of a number of rf waves traveling in opposite directions around the magnetron periphery causing the time varying fields to exist in the space between anode and cathode. This space is called the interaction space because it is the space common to the rf fields and the electrons where interaction takes place. Only one type of wave is considered of importance in the present discussion. This is a wave traveling in the same direction as the electrons moving around the interaction space so that interaction between the electrons and the wave is appreciable. Waves traveling in the opposite direction are considered as contributing only to higher order corrections.

EQUATIONS GOVERNING SPACE CHARGE BEHAVIOR

The mathematical analysis of space charge behavior in a magnetron is based on the force equation for an electron in an electromagnetic field. This equation in vector form is the following

$$d\mathbf{v}/dt=-e/m(\mathbf{F}+\mathbf{v}\times\mathbf{B}), \tag{1}$$

where e=absolute value of electronic charge, m=mass of electron, $\mathbf{v}$=velocity of electron, $\mathbf{F}$=vector value of electric field, $\mathbf{B}$=vector value of magnetic field.

The magnetron will be considered with axial alignment in the z-direction. In this case it is convenient to resolve the vector quantities in Eq. (1) into components in cylindrical coordinates. If proper account is taken of the symmetries in the magnetron, Eq. (1) may be written as follows:

$$\frac{d\mathbf{v}}{dt}=\mathbf{r}_1\left\{\frac{dv_r}{dt}-r\omega^2\right\}+\boldsymbol{\theta}_1\left\{2v_r\omega+r\frac{d\omega}{dt}\right\}$$
$$=-e/m(\mathbf{r}_1F_r+\boldsymbol{\theta}_1F_\theta+\mathbf{r}_1r\omega B_z-\boldsymbol{\theta}_1v_rB_z). \tag{1a}$$

Here the quantities with subscript (1) are unit vectors. $\mathbf{v}$ has been replaced by

$$\mathbf{v}=\mathbf{r}_1v_r+\boldsymbol{\theta}_1r\omega. \tag{2}$$

$r\omega$ is therefore the magnitude of the θ-directed component of velocity. ω is the angular velocity of the electron. The components F_z, v_z, B_θ, B_r are considered non-existent as would be the case in an ideal magnetron. By separation of the r- and θ-directed components in Eq. (1a) we obtain the following two important equations for scalar quantities.

$$dv_r/dt-r\omega^2=-e/m(F_r+r\omega B_z), \tag{3a}$$

$$2v_r\omega+rd\omega/dt=-e/m(F_\theta-v_rB_z). \tag{3b}$$

There are several results to be derived from these two equations based on various assumptions which may be used. We will consider two cases of interest. Since B is always in the z-direction we will drop the subscript z.

THE STATIC MAGNETRON, NO rf VOLTAGE PRESENT

This case will be discussed very briefly to give the necessary results, since it has been treated in the references.

We will first assume that no rf field is present; therefore, $F_\theta=0$. In this case Eq. (3b) can be written

$$2dr/dt(\omega_L-\omega)=rd\omega/dt. \tag{4}$$

v_r has been replaced by dr/dt and $eB/2m$ has been replaced by ω_L. The equation is readily solved by separa-

*The material in this paper is based on work done for the U. S. Army Signal Corps on Contract No. W36-039 sc-35561.

[1] L. Brillouin, Phys. Rev. **60**, 385–396 (1941). J. P. Blewett and S. Ramo, Phys. Rev. **57**, 635–641 (1940). J. C. Slater, *Microwave Electronics* (D. Van Nostrand Company, Inc., New York, 1950), Chapter VIII. L. Page and N. I. Adams, Phys. Rev. **69**, 492 (1946). G. B. Collins, *Microwave Magnetrons* (McGraw-Hill Book Company, Inc., New York, 1948), Part II.

tion of variables and assuming ω to be zero at the boundary of the cathode r_c.

$$\omega=\omega_L[1-(r_c^2/r^2)]. \tag{5}$$

This result may be substituted back into Eq. (3a) yielding an equation which can be solved simply only under the restrictive assumption that

$$dv_r/dt=0.$$

Now,

$$dv_r/dt=\partial v_r/\partial t+(\partial v_r/\partial r)v_r; \tag{6}$$

therefore, in order for dv_r/dt to be zero,

$$\partial v_r/\partial t=0,$$

implying a steady-state condition, and

$$v_r=0$$

implying that no radial current exists.

The potential distribution resulting in this case is actually obvious from Eq. (5) and the assumptions just stated

$$E-E_c=1(m/2e)\omega_L^2r^2[1-(r_c^2/r^2)]^2 \tag{7}$$

where E=potential at r and E_c=potential at r_c. This relationship states simply that the energy taken by the electron from the electric field is equal to its kinetic energy. Using the poisson equation the space charge density under these conditions can be shown to be

$$\rho/\epsilon_0=-2m/e\omega_L^2[1+(r_c^4/r^4)] \tag{8}$$

where ϵ_0=dielectric constant of free space$=1/36\pi\times10^{-9}$ farad/m. The electric field is given by

$$F_r=-\partial E/\partial r=(m/e)r\omega_L^2[(r_c^4/r^4)-1]. \tag{9}$$

It is important to realize that Eqs. (7)–(9) represent distributions *within* the space charge swarm only for the special case of no radial current throughout the swarm. Equations (7) and (9) do hold more significance for the swarm boundary. Such a boundary must exist if the anode potential is less than the value necessary to make the swarm extend all the way out to the anode. The potential is given by substituting the anode radius (r_a) into Eq. (7). [The electron must have *at least* the energy corresponding to the angular velocity of Eq. (5) since this result is independent of the solution of Eq. (3a).] We are led to the conclusion that if the swarm is bounded within the interaction space between anode and cathode no current is crossing the boundary, therefore Eqs. (7) and (9) must hold at the boundary. By the gaussian theorem the total space charge per unit length within a boundary at radius r_H will, therefore, be

$$\begin{aligned}\tau_H&=2\pi\epsilon_0[-(\partial E/\partial r)] \text{ at } r_H\\&=(1\pi/2)\epsilon_0(e/m)B^2r_H^2[(r_c^2/r_H^2)-1].\end{aligned} \tag{10}$$

We must conclude that the total space charge within a bounded swarm is independent of the potential, field and space charge distribution within the swarm. Electrons may be circulating around the cathode or streaming radially outward and inward through the swarm provided they do not cross the boundary r_H. If there are radial velocities within the swarm the potential within the swarm must be greater than that given by Eq. (7).

The purpose of this discussion is to emphasize the fact that the relationship for the radius of the swarm boundary in terms of anode potential which follows is not dependent on the restrictive assumptions corresponding to the existence of the distributions of Eqs. (7)–(9) throughout the swarm. It is only dependent on the assumptions that the initial velocity and the radial velocity are zero and that no sources of energy exist other than the dc field.

If we consider the magnetron as a cylindrical diode containing the space charge swarm just described and having no gradient at the cathode then

$$E_a-E_H=-(\tau_H/2\pi\epsilon_0)\log(r_a/r_H),$$

where E_a=anode potential and E_H=value given by Eq. (7) for r_H. We have the complete result

$$E_a=B^2\frac{e}{8m}r_H^2\left[2\left(1-\frac{r_c^4}{r_H^4}\right)\ln\frac{r_a}{r_H}+\left(1-\frac{r_c^2}{r_H^2}\right)^2\right]. \tag{11}$$

CONDITIONS FOR SYNCHRONISM WITH THE rf FIELD IN THE MAGNETRON

In the discussion of this section it is not actually necessary for an rf field to be present. It is only necessary to define the angular velocity around the interaction space which corresponds to a particular rf field which might be present.

It will be convenient to define some terms which will simplify the following discussion and will make the terminology the same as that generally used in discussing the scaling of magnetrons.[2] These definitions also have the advantage of making most of the equations dimensionless. The presence of the rf is described in the usual manner.[2] The rf wave with which the electrons interact is considered as traveling around the interaction space between anode and cathode, with an angular velocity given by

$$\omega_n=\frac{2\pi f}{n} \text{ radians/sec} \tag{12}$$

where f is the frequency of the rf impressed on the magnetron anodes and n is the mode number, equal to $N/2$ in the π-mode where N is the number of anodes.

The kinetic energy of an electron at the anode having the angular velocity given by Eq. (12) defines a voltage which we will call E_0

$$E_0=(m/2e)\omega_n^2r_a^2. \tag{13}$$

The magnetic field which would allow this angular velocity to exist at the anode of a non-oscillating magnetron is defined by Eq. (5) when ω is equal to ω_n

[2] See reference 1, J. C. Slater, Chapter 13.

and r set equal to r_a

$$B_0=\frac{2m}{e}\omega_n\frac{1}{[1-(r_c^2/r_a^2)]}. \tag{14}$$

In terms of these variables Eq. (7) may be written

$$(E-E_c)/E_0=(B/B_0)^2. \tag{7a}$$

E_0 and B_0 depend only on dimensions of the magnetron, the frequency impressed on the magnetron anodes and physical constants.

At a particular voltage, outer electrons in the swarm reach the angular velocity of the rf field given by Eq. (12). This we will call the synchronous voltage. In order to get an idea of the radius of the space charge for which synchronism takes place, it is convenient to rewrite Eq. (11) in terms of the new variables as follows:

$$\frac{E_a}{E_0}=\left(\frac{B}{B_0}\right)^2\left(\frac{r_H}{r_a}\right)^2\left\{2\frac{1-(r_c^4/r_H^4)}{[1-(r_c^2/r_a^2)]^2}\ln\frac{r_a}{r_H}+\left[\frac{1-(r_c^2/r_H^2)}{1-(r_c^2/r_a^2)}\right]^2\right\}. \tag{11a}$$

Using the definition of B_0 we write

$$\frac{\omega}{\omega_n}=\frac{B}{B_0}\frac{1-(r_c^2/r^2)}{1-(r_c^2/r_a^2)}. \tag{15}$$

When $\omega=\omega_n$

$$\frac{B_n}{B_0}=\frac{1-(r_c^2/r_a^2)}{1-(r_c^2/r_n^2)}. \tag{16}$$

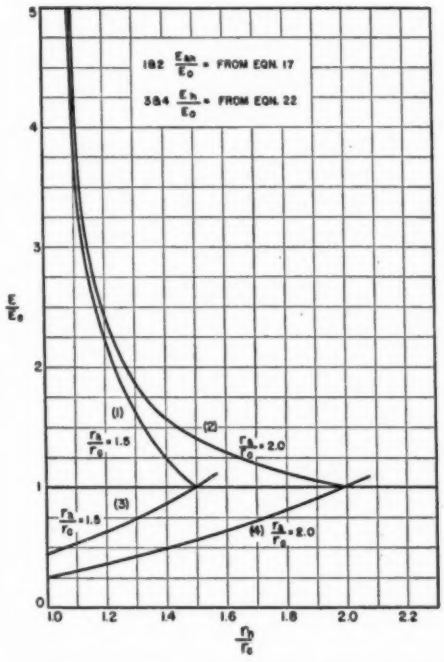


FIG. 1. Anode voltage and space charge cloud boundary voltage as function of radius of space charge cloud boundary for synchronism of outer electrons.

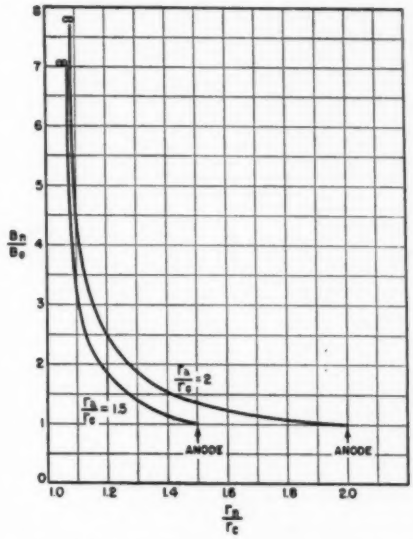


FIG. 2. Magnetic field as function of radius of space charge cloud boundary for synchronism of outer electrons. [From Eq. (16).]

The subscript n denotes values for which the angular velocity of the outer edge of the electron swarm reaches synchronism with the angular velocity of the wave. The voltage for which synchronism occurs at a given radius is given by substitution of B_n/B_0 from Eq. (16) into Eq. (11). We are discussing the outer edge of the cloud so $r_H=r_n$.

$$\frac{E_{an}}{E_0}=\frac{r_n^2}{r_a^2}\left[2\frac{1+(r_c^2/r_n^2)}{1-(r_c^2/r_n^2)}\ln\frac{r_a}{r_n}+1\right]. \tag{17}$$

An interesting conclusion to be drawn from Eqs. (16) and (17) is that for given values of ω_n and r_n and a given tube geometry there is only one set of values for E_a/E_0 and B/B_0, which will allow the edge of the electron swarm to rotate in synchronism. Equations (16) and (17) are plotted in Figs. 1 and 2 for two values of r_a/r_c.

By eliminating r_n between these two equations, we may write an equation relating E_{an}/E_0 and B_n/B_0 for the condition that the outermost electrons in the cloud have just reached synchronism.

$$\frac{E_{an}}{E_0}=\frac{B_n}{B_0}\frac{r_c^2/r_a^2}{(B_n/B_0)-[1-(r_c^2/r_a^2)]}\times\left\{2\left[\frac{B_n}{B_0}\frac{2}{1-(r_c^2/r_a^2)}-1\right]\times\log\frac{\{(B_n/B_0)-[1-(r_c^2/r_a^2)]\}^{\frac{1}{2}}}{r_c/r_a(B_n/B_0)^{\frac{1}{2}}}+1\right\}^3. \tag{18}$$

[3] See H. W. Welch, Proc. Inst. Radio Engrs. p. 1434 (December, 1950) for experimental confirmation of this expression.

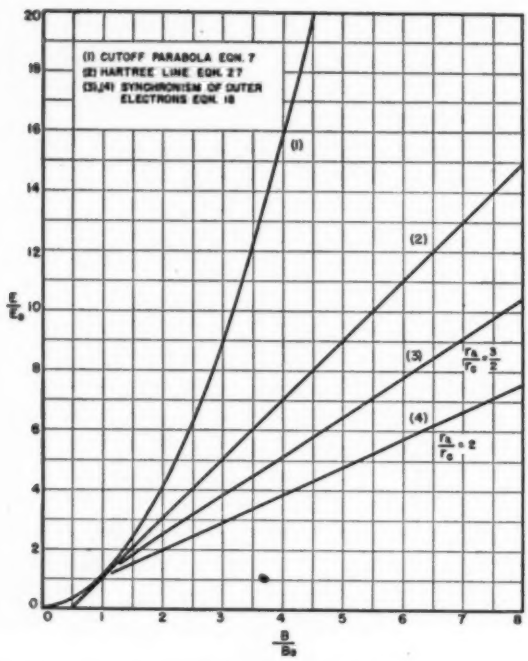


FIG. 3. Anode voltages for cutoff, Hartree, and synchronous boundaries as function of magnetic field.

SYNCHRONISM: CONDITION FOR INITIATION OF OSCILLATION

It is convenient at this point to discuss an idealized type of space charge behavior which is similar in principle to the case in the static magnetron corresponding to distributions of Eqs. (7)–(9) existing throughout the space charge swarm. It will be assumed that an rf field is present traveling around the interaction space with an angular velocity given by Eq. (12). At the voltage given by Eq. (18) electrons at the outer edge of the swarm formed by the dc fields attain this velocity. Above this voltage the outer electrons are trying to travel faster than the field. The field has tangential components which may either oppose or aid the motion of the electron, depending on the position of the electron relative to the traveling wave. The electrons which take energy from the field will drift inward and eventually give up their excess energy to the cathode. In regions of opposing field the electrons will give up energy to the rf field and be slowed to synchronous velocity. During the process of energy absorption by the field the electrons will drift outward from the edge of the static swarm due to a decrease in the Lorentz force. The potential distribution outside the swarm will thus be altered by presence of the space charge.

Let us now examine the potential distribution in a region where electrons are assumed to be in synchronism. This can be derived quite simply from the original force equation if we make the initial assumption that the angular velocity of the electrons is the synchronous velocity ω_n. In this case, Eq. (3a) is written

$$dv_r/dt - r\omega_n^2 = -e/m(F_r + r\omega_n B_z).$$

We assume as we did before steady state with no radial motion; i.e., all energy is kinetic energy of rotation. In this case it will be convenient to make the substitution:

$$\omega_c = Be/m = 2\omega_L. \tag{19}$$

We have therefore

$$F_r = -\partial E/\partial r = (m/e)r\omega_n(\omega_n - \omega_c). \tag{20}$$

This equation can be rewritten

$$-\frac{\partial}{\partial r}\left(\frac{E}{E_0}\right) = 2\frac{r}{r_a^2}\left(1 - \frac{\omega_c}{\omega_n}\right).$$

Integrating this becomes:

$$-E/E_0 = [1 - (\omega_c/\omega_n)](r^2/r_a^2) + C. \tag{21}$$

In order to evaluate C we will make use of the assumption that all energy is kinetic energy of angular velocity. Thus as the synchronous boundary $r = r_n$

$$eE_n = m/2\omega_n^2 r_n^2$$

and at the anode

$$eE_0 = m/2\omega_n^2 r_a^2.$$

Therefore, dividing one equation by the other, we have at $r = r_n$

$$E_n/E_0 = r_n^2/r_a^2. \tag{22}$$

This equation is plotted for comparison with Eq. (17) in Fig. 1. Using this relationship as a boundary condition to evaluate C, Eq. (21) becomes

$$\frac{E}{E_0} = -\frac{r^2}{r_a^2} + 2\frac{r_n^2}{r_a^2} + \frac{\omega_c}{\omega_n}\left(\frac{r^2}{r_a^2} - \frac{r_n^2}{r_a^2}\right). \tag{23}$$

There are two ways of expressing ω_c/ω_n which are useful in simplifying Eq. (23). These are

$$\frac{\omega_c}{\omega_n} = \frac{B}{B_0}\frac{1}{1 - (r_c^2/r_a^2)} \tag{24}$$

[from the definitions of ω_c in (19) and B_0 in (14)] and

$$\omega_n/\omega_c = \tfrac{1}{2}[1 - (r_c^2/r_n^2)]. \tag{25}$$

The latter relationship is based on the assumption that the solution in Eq. (5) holds out to r_n. Rewrite (23)

$$\frac{E}{E_0} = \frac{\omega_c}{\omega_n}\left(\frac{r^2}{r_a^2} - \frac{r_n^2}{r_a^2} + 2\frac{r_n^2}{r_a^2}\frac{\omega_n}{\omega_c}\right) - \frac{r^2}{r_a^2}.$$

If we make the substitutions (24) and (25) for the quantity ω_c/ω_n outside the bracket and inside the bracket, respectively, the result becomes

$$\frac{E}{E_0} = 2\frac{B}{B_0}\frac{(r^2/r_a^2) - (r_c^2/r_a^2)}{1 - (r_c^2/r_a^2)} - \frac{r^2}{r_a^2}. \tag{26}$$

This new potential distribution exists throughout the distribution of space charge which forms outside the static magnetron space charge cloud because of electrons

being slowed to synchronism. If the outermost electrons of this cloud are just capable of reaching the anode the magnetron will draw plate current. The electrons must give up energy to reach this radius; therefore, as soon as the magnetron is capable of drawing current, it will be able to use this energy to generate oscillations. This assumes of course that ω_n corresponds to a frequency at which the magnetron resonator will oscillate. This condition may be imposed upon Eq. (26) and we have, for $r=r_a$

$$E_a/E_0=(2B/B_0)-1. \tag{27}$$

This is a well known relationship obtained in a different way by Hartree[4] and defining the voltage at which oscillations begin. The assumptions and method used here make clear that the potential of Eq. (27) with the distribution of Eqs. (7) and (26) represent the minimum energy of the individual electron which allows electrons to run in circular orbits extending all the way out to the anode and having synchronous velocity at the anode.

If the anode potential is raised above the value given by Eq. (27) the electrons must have more energy and therefore radial current will exist or the synchronous angular velocity must be increased.

Equations (7), (18), and (27) are plotted for comparison in Fig. 3 in dimensionless variables. It should be noted that the synchronous voltage of Eq. (18) is reached well below the Hartree voltage of Eq. (27). This means that, if a radio frequency is impressed on the anodes the synchronous space charge swarm should start to form well below the Hartree voltage. The synchronous voltage can be observed by measuring the resonant wavelength of a hot magnetron as the anode voltage is raised. When the synchronous voltage is reached a rather abrupt increase in resonant wavelength is observed due to the expansion of the synchronous space charge.[5]

SYNCHRONOUS SPACE CHARGE

One other equation of some interest can be derived from these relationships. This is the equation for the space charge under the conditions of what we will henceforth call the synchronous space charge distribution. If we apply the poisson equation to Eq. (20) the result is

$$\rho=-2\epsilon_0 m/e\omega_n(\omega_c-\omega_n). \tag{28}$$

The space charge in the synchronous space charge distribution is therefore a constant independent of the radius and dependent only on the angular velocity of the wave, the magnetic field and physical constants.

If the distribution in the portion of the space charge cloud traveling at less than the synchronous velocity ω_n is presumed to be the static magnetron distribution given by Eq. (8); the space charge density at the outside edge of the static magnetron swarm and the inside edge of the synchronous cloud are of different magnitude. Making the substitution

$$\omega_n=\omega_L[1-(r_c^2/r_n^2)]$$

in the equation for the static magnetron space charge distribution (8), this equation becomes

$$\rho=-2\epsilon_0(m/e)[(\omega_c^2/2)-\omega_n(\omega_c-\omega_n)] \tag{8a}$$

at the boundary between the two distributions. Thus, at the boundary

$$\frac{\rho_{\text{static}}}{\rho_{\text{synch}}}=-\frac{1}{2}\frac{\omega_c^2}{\omega_n(\omega_c-\omega_n)}-1. \tag{29}$$

Similar algebraic manipulation with the equations for the field and for the potential distribution will show that these quantities are exactly equal at the boundary between the two clouds.

In order to summarize the two types of distributions they are plotted for comparison in Fig. 4. Figure 5 is a pictorial representation of the space charge showing the inner static magnetron swarm and the outer synchronous swarm. Since electrons can only move outward in regions of decelerating rf field the outer swarm is shown as consisting of "spokes" of space charge. This picture of space charge spokes in the oscillating magnetron has been presented by Hartree, Slater, and others and is generally accepted by workers in the magnetron field. Equations (16), (17), (18), and (28) represent relationships which make possible quantitative calculation of synchronous space charge or "spoke" behavior.

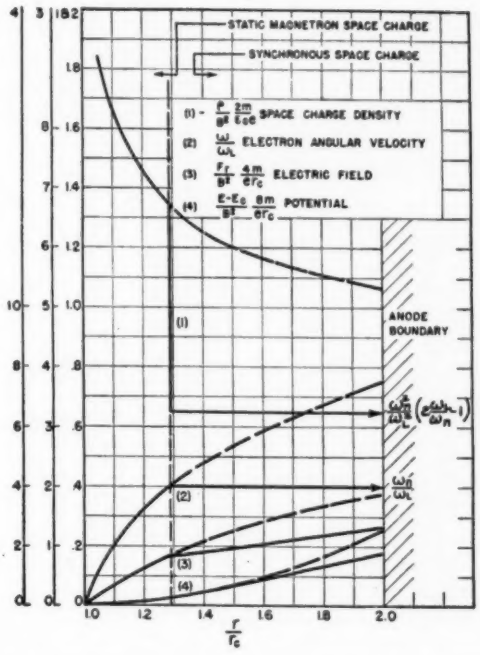


FIG. 4. Distribution in typical magnetron as synchronous swarm reaches the anode for $\omega_n/\omega_L=0.4$. Note: dotted curves are continuation of static distribution.

[4] D. R. Hartree, "Estimates of Electron Energies, Oscillation Amplitudes, and Efficiencies in a Magnetron Operating under Rotating Space Charge Conditions," C. V. D. Report, Ref. Mag. 11.

[5] Experiments of this type are discussed in reference 3.

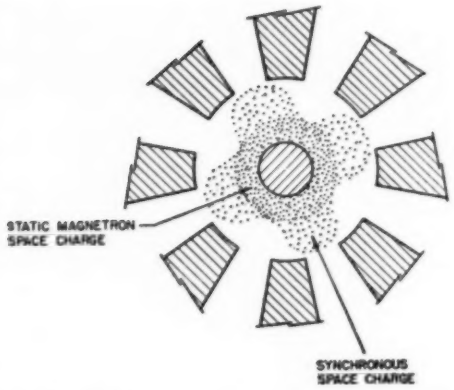


FIG. 5. Pictorial representation of space charge within interaction space.

INITIATION OF ANODE CURRENT; NOISE IN MAGNETRONS[6]

The main purpose in presenting completely the preceding analysis of magnetron synchronous space charge is to make clear the nature of the assumptions which are used and therefore the conditions under which the analysis is true. The resulting values of voltages for the magnetic diode cutoff [Eq. (7) for $r=r_a$] and voltage for initiation of oscillation [Eq. (27)] have been rightfully criticized for not giving good agreement with experimentally observed results. However, examination of the assumptions shows that both of these voltages are based on the condition that the electron has acquired the minimum amount of kinetic energy of angular velocity required for existence in a stable state at the anode radius. Thus plate current must exist at higher voltages. If current initiates at lower voltages, the electrons must have a smaller minimum kinetic energy and thus smaller angular velocities. If this is the case, they must have been slowed down by some mechanism in their transit to the anode. The only significant mechanism by which the electrons can be decelerated is in a transfer of energy from the dc field to some electromagnetic field. In the case of the initiation of oscillation, it is assumed that electrons have been slowed to synchronous angular velocities by the interaction with the oscillatory electromagnetic field. However, it is experimentally observed that noise voltages exist in pre-oscillating magnetrons and in static magnetrons operating below cutoff, and it is logical to assume that the production of these voltages is the mechanism by which electrons give up their energy, and therefore reach the anode at lower than predicted values of dc voltage. Conversely, if noise is produced at a higher level than accounted for by the cathode temperature, it is necessary that anode current be drawn. This is true for the simple reason that, if noise energy is to be produced by the system, then energy must be given the system by the power supply. The noise probably originates in the cathode as thermal noise, etc., and is amplified by the process of continuous interchange of energy between the dc field and the noise field through the medium of the electron swarm. On the basis of these ideas, it is suggested that experiments designed to cast light on the origins of noise produced by magnetrons should be made on non-oscillating structures and correlated carefully with the dc current drawn by the magnetron. It is also possible that an amplifier could be constructed employing the principle of interaction between layers of magnetron space charge traveling at different angular velocities similar to the already constructed electron-wave tube.[7]

[6] R. L. Sproull, J. Appl. Phys. **18**, 314–320 (1947): V. Mayper, "Noise Generation in Pre-Oscillating Magnetron," Quarterly Progress Report, M.I.T. Research Laboratory of Electronics, 4–6 (January 15, 1948).

OSCILLATION; ELECTRONIC EFFICIENCY

The definition of the voltage for initiation of oscillation given by Eq. (27) has a similarity in principle to the magnetic diode cutoff voltage in that the electrons have the minimum kinetic energy of angular velocity required for their existence at the anode. It is noteworthy that the kinetic energy of the electrons at the anode is in general less than the energy they have acquired, the difference being the energy supposedly delivered to a rotating electromagnetic field with which they are synchronous. Thus it is possible to calculate the maximum electronic efficiency which is obtainable under the given conditions of voltage and tube geometry. If E_a is the anode voltage at which anode current is drawn, representing energy delivered to the electron, and E_0 is the voltage defined by Eq. (15), representing the kinetic energy of the synchronous electron, then maximum electronic efficiency is given by

$$\eta_e=(E/E_0-1)/E/E_0. \tag{30}$$

Thus for interaction efficiency greater than 80 percent, E/E_0 must be greater than 5. This places certain restrictions on r_a and n and determines B/B_0 through Eq. (27).

A further consideration which must be included in an exact analysis of the oscillating magnetron is the effect of loading. The fact that an electron may exist at the anode and thus permit oscillation to begin, does not imply that the negative conductance which the space charge provides will meet the conditions imposed by the load. Thus, experimentally it is observed that voltage for initiation of oscillation depends on the loading in addition to the other factors. The maximum electronic efficiency given by Eq. (30) is observed in some cases, but can only exist under optimum load conditions. No analysis has been presented, to our knowledge, which predicts quantitatively the effect of loading conditions on efficiency, although experimental data on a particular tube enables one to make semiquantitative predictions.

[7] A. V. Haeff, Proc. Inst. Radio Engrs. (January, 1949).

The Deformation of Gold Wire at High Temperature*

B. H. ALEXANDER, M. H. DAWSON, AND H. P. KLING
Metallurgical Laboratories, Sylvania Electric Products Inc., Bayside, New York

(Received November 15, 1950)

Gold wires were subjected to small tensile stresses at high temperatures and the strain-time relationships were determined. These creep curves were similar to those generally obtained for metals at lower temperatures, except that the minimum creep rate was proportional to the applied stress; i.e., the flow was viscous. This coefficient of viscosity was determined as well as its variation with temperature. The surface energy of solid gold was also determined. Theories that have been presented to explain the viscous behavior of metals are briefly discussed.

INTRODUCTION

LITTLE is known about the mechanical behavior of metals when subjected to small stresses at temperatures near their melting points. Such knowledge would be helpful in understanding the mechanisms of sintering and of creep, especially microcreep, of metals. It is of particular interest to know whether metals may behave viscously† under certain circumstances. Cottrell and Jaswon,[1] Nabarro,[2] and Herring[3] have considered the latter point theoretically and have proposed mechanisms which would allow viscous flow in metals.

Using small stresses, Chalmers[4] found that, at room temperature, the initial strain rate of tin was directly proportional to the stress. These experiments were thus indicative of a type of flow which could be described as viscous for a small range of strains. Udin, Shaler, and Wulff[5] measured the strains in copper wires under small stresses after heating them several hours at high temperatures. They assumed that the wires behaved viscously and were able to determine the load which was necessary to counteract the surface tension, and thus the latter could be evaluated. The present work was designed to determine the strain-time curves for gold wires under small tensile stresses and over a temperature range near the melting point of gold. The existence of a purely viscous type of behavior should be detectable by such experiments.

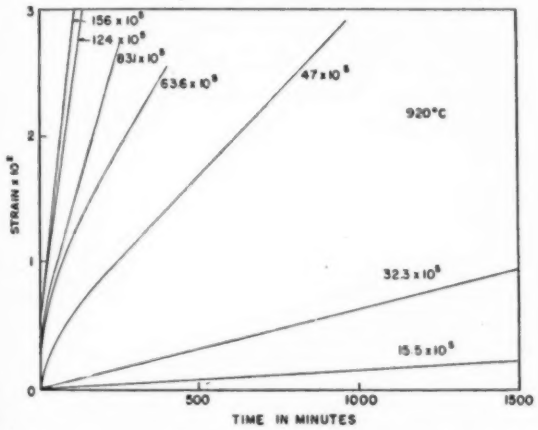


FIG. 1. Strain *versus* time curves for 1-mil gold wire at 920°C. The curves are identified by the stress in dynes/cm².

EXPERIMENTAL

Gold wire (999 fine) with a nominal diameter of 0.001 in. was used in all experiments. The diameter of the wire as determined by weighing measured lengths was found to be 0.0028 cm. This latter figure was further verified by direct microscopic measurements, and in all further calculations this figure was used. The specimens of this wire used for testing were about 12 inches long, and knots were tied near the center in order to mark the gauge length, which was usually from 10 to 12 cm. These specimens were suspended in heating elements and small tungsten weights were tied to the bottom of the specimens in order to achieve the desired stresses. The heating elements consisted of thin walled nickel tubes ($\frac{1}{4}$-in. inside diameter) which were slotted vertically to permit observation of the gauge markers. The nickel heating elements were in turn mounted inside of bell jars which were partially evacuated (~0.1 mm pressure) during the experiments.

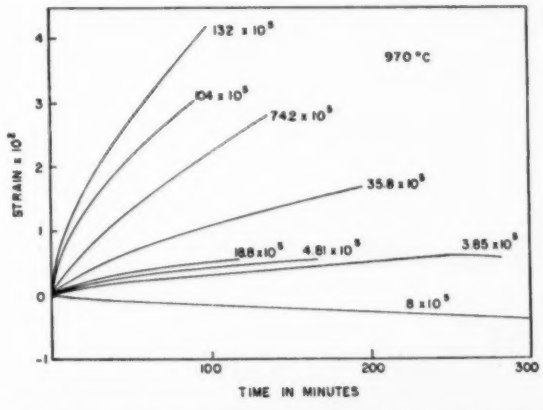


FIG. 2. Strain *versus* time curves for 1-mil gold wire at 970°C. The curves are identified by the stress in dynes/cm².

* This work was supported in part by AEC Contract AT-30-1 GEN 367.

† By viscous behavior we mean only that the strain rate is directly proportional to the applied stress.

[1] A. H. Cottrell and M. A. Jaswon, Proc. Roy. Soc. (London) A199, 104–14 (1949).
[2] F. R. N. Nabarro, Report of a Conference on the Strength of Solids, The Physical Society of London (1948).
[3] C. Herring, J. Appl. Phys. 21, 437 (1950).
[4] B. Chalmers, J. Inst. Metals 61, 103 (1937).
[5] Udin, Shaler, and Wulff, J. Metals 1, 186–90 (1949).

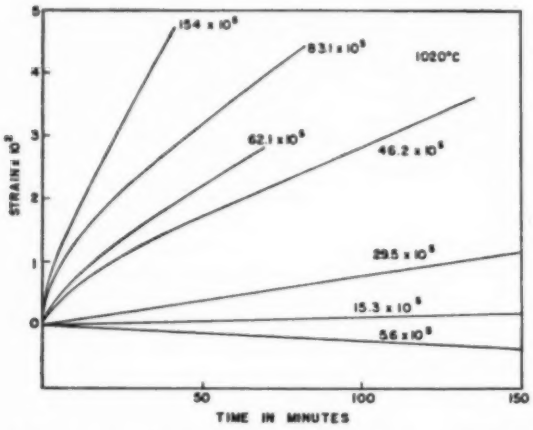


FIG. 3. Strain *versus* time curves for 1-mil gold wire at 1020°C. The curves are identified by the stress in dynes/cm².

The temperature was measured by means of thermocouples welded to the outer surface of the nickel heating element. By suspending a thermocouple simulating one of the gold creep specimens, it was found that the specimen itself was at a temperature 80°C lower than the heating element. This was confirmed by raising the temperature of the heating element until a gold specimen hanging inside would melt; melting only occurred when the heating element was 80°C above the melting point of gold. A constant-voltage regulator was used as a power supply to the heating elements, and frequent checks were made during each experiment in order to maintain a constant temperature. It is felt that the temperature did not vary more than 3°C during an experiment, although the method of determining the true temperature allowed a possible error of 5°C.

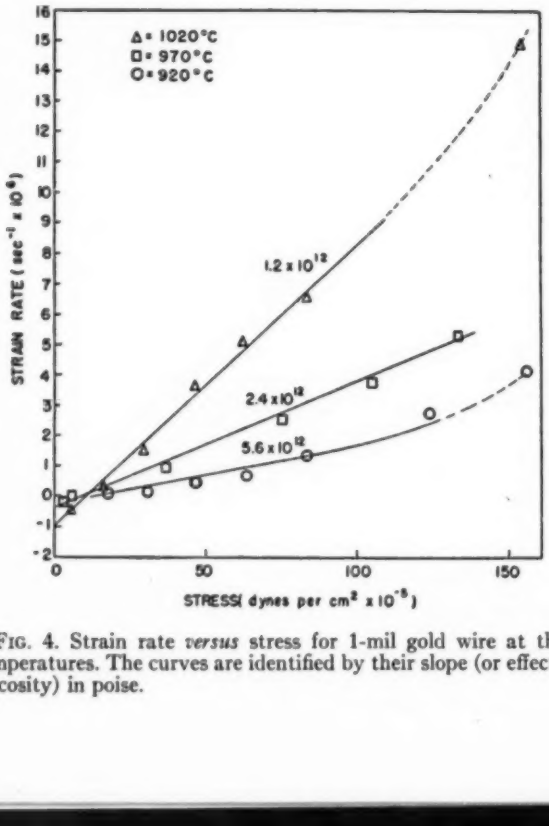


FIG. 4. Strain rate *versus* stress for 1-mil gold wire at three temperatures. The curves are identified by their slope (or effective viscosity) in poise.

The strain was measured by a traveling microscope which could be read to the nearest 0.001 cm, although the true accuracy was probably near 0.01 cm. Using this latter figure, with a gauge length of 10 cm, we conclude that the actual strain was measured to a little better than 0.1 percent. The total elongation or contraction of a specimen was between 1 and 5 percent, depending on the stress. All specimens were annealed for 10 minutes at 750°C while under load in order to remove any bowing. The specimens were then cooled to room temperature, measured, and heated immediately to the testing temperature. It took only a few seconds to heat the specimens to the desired temperature; from then on, the specimen was measured as frequently as necessary to follow the strain accurately.

DISCUSSION OF RESULTS

The strain-time curves obtained for the three temperatures, 920°, 970°, and 1020°C, are shown in Figs. 1 through 3, respectively. It can be seen that at the lower

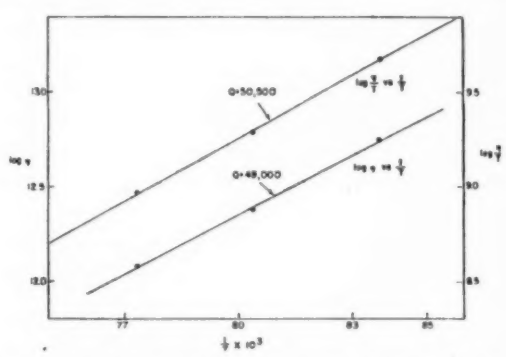


FIG. 5. Plots showing the variation of the log of η (viscosity coefficient) and η/T with the reciprocal of the absolute temperature, T. The slope Q is given in calories per mole.

stresses these curves are linear; at higher stresses the curves show an initial rapid elongation followed by a slower approximately constant rate of extension. At stresses so low that the elongation was negative, the curves are also linear throughout, although the behavior here was somewhat erratic. Some specimens showed a tendency to reach a constant length after some shrinkage; others would alternately stretch and shrink and finally break after an over-all shortening; still others would shorten continually at a constant rate and would finally break. The curves shown for the lower stresses represent the average behavior of several samples.

The slopes of the straight-line portion of the strain-time curves are plotted against the corresponding stresses in Fig. 4. It can be seen that these plots can be satisfactorily represented by straight lines and thus indicate that the type of flow predominating in these experiments can be described as effectively viscous inasmuch as the strain rate is directly proportional to the stress. At stresses greater than about 120×10^5 dynes/cm², the strain rate is no longer directly proportional to the stress.

We may use the curves shown in Fig. 4 to calculate the surface tension if we assume that the applied stress which causes a zero strain rate is just sufficient to counteract the stress due to surface tension. This gives the following values for the surface tension:

Temperature °C	Surface tension (dynes/cm)
920	1680
970	1280
1020	1400

About the best we can conclude from these results is that the surface tension is in the range of 1200 to 1700 dynes/cm in the temperature range 920° to 1020°C. It is possible that more accurate strain measurements and more numerous experiments at the lower stresses would give more accurate values for the surface tension. However, as we shall see later, the manner in which the wire deforms at these low stresses would seem to militate against this.

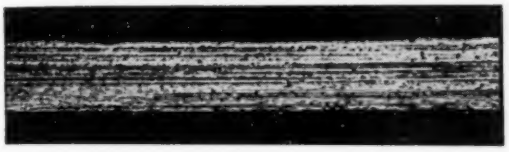


FIG. 6. Longitudinal section of 1-mil gold wire as drawn. ×700.

FIG. 7. Longitudinal section of 1-mil gold wire after 96 hours at 1020°C under a stress of 8 psi. ×700. This and the following two figures show "kinking" of the wire at grain boundaries.

The effective viscosity coefficients, taken as the slopes of the lines in Fig. 4 are shown in Fig. 5, where the logarithm of the viscosity coefficient is plotted against the reciprocal of the absolute temperature. The activation energy calculated from this plot is 48,000 cal/M, which is in reasonable agreement with the activation energy for self-diffusion in Au, which is reported to be 51,000 cal/M.[6] We may conclude, for the moment, that the viscous behavior of the gold in the experiments described above is associated with self-diffusion.

Three theoretical explanations have been proposed which can be used to explain this type of flow in metals. Nabarro[2] made the assumption that vacancies can migrate to and from the boundaries of mosaic blocks in a metal under the influence of an applied stress. With this assumption he obtains the following expression for the viscosity coefficient at low stresses:

$$\eta \cong A^2kT/Da^3, \qquad (1)$$

[6] H. A. C. McKay, Trans. Faraday Soc. 34, 845 (1938).

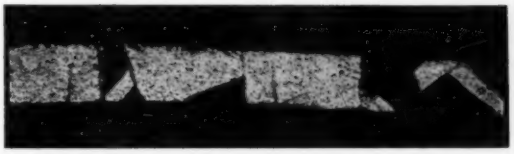


FIG. 8. After 24 hours at 970°C under a stress of 25 psi. ×700.

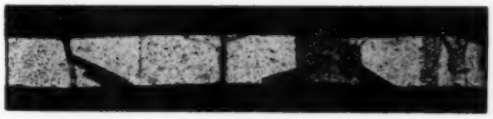


FIG. 9. After 25 minutes at 920°C under a stress of 10 psi. ×700.

where η = viscosity coefficient (dynes-sec/cm²), A = dimension of mosaic block, k = Boltzmann's constant, D = self-diffusion coefficient, a^3 = atomic volume, and T = absolute temperature.

Nabarro's derivation has been refined by Herring[3] who gives two formulas for the effective viscosity coefficient, depending on whether the tangential stresses at the mosaic boundaries are relaxed or not. These formulas are

$$\eta = A^2kT/4Da^3 \text{ (for tangential stresses not relaxed)}, \qquad (2)$$

and

$$\eta = A^2kT/10Da^3 \text{ (for tangential stresses relaxed)}. \qquad (3)$$

FIG. 10. After 2½ hours at 920°C under a stress of 150 psi. ×700. The offsetting at the grain boundary as seen here was only observed at higher stresses.

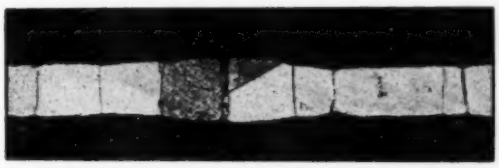
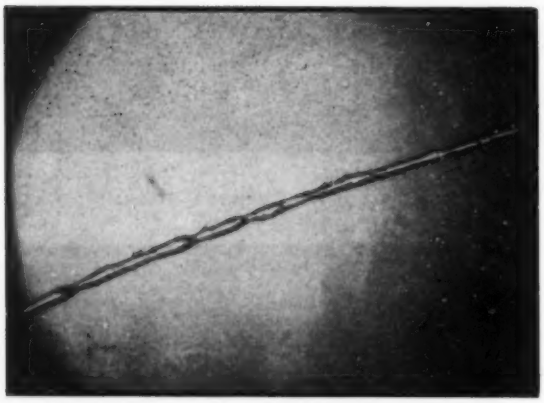
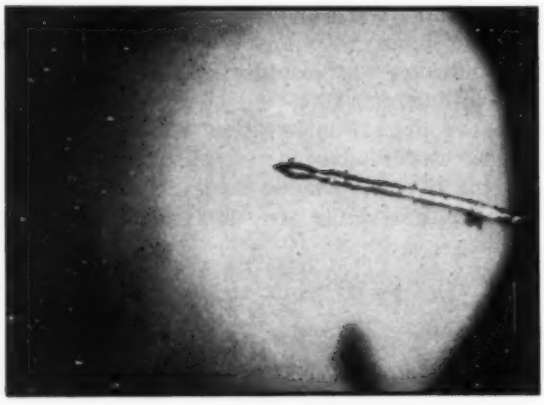


FIG. 11. After 100 hours at 1020°C under a stress of 10 psi. ×700. The development of the bulges and waists shown in this and the following figure was characteristic of the wires that had contracted slightly.

FIG. 12. Same as Fig. 11, another section.





FIGS. 13 and 14. 0.4-mil gold wire after 120 hours at 1055°C. ×100.

Herring[3] has also made a calculation of the effective viscosity coefficient under the assumption that vacancies migrate to and from grain boundaries under the influence of an applied stress. For the present case his formula is

$$\eta = 2kTRL/3BDa^3, \qquad (4)$$

where R = radius of the wire, L = length of the grains in the wire, and B = a constant depending on the ratio R/L.

A third possible mechanism of flow, which might account for the present observations, is the motion of dislocations with an atmosphere of impurity atoms, as described by Cottrell and Jaswon.[1] A direct calculation of the viscosity coefficient according to this theory cannot be made because of the lack of knowledge about the number of dislocations and the forces necessary to move them away from their atmosphere.

All three of these theories would predict that the viscosity coefficient should vary with temperature to the same degree that the diffusion coefficient does; this is borne out by the plots shown in Fig. 5. It will be noticed that, according to the formulas given by Nabarro and Herring, it is more proper to plot the logarithm of η/T *vs* $1/T$ in order to obtain the same activation energy as for diffusion. Such a plot is shown in Fig. 5, and it can be seen that, although this does not greatly affect the activation energy, it gives a value of 50,000 cal/M, which is in better agreement with that obtained by diffusion measurements (51,000 cal/M).

The ratio of expression (4) to (2) is $8RL/3A^2B$. This ratio is about 86 for the gold wires used in this investigation, assuming that A is about 10^{-4} cm. It should, therefore, be possible to decide whether diffusion to and from mosaic boundaries or grain boundaries is the operative process, if all the constants in the formulas are known with even moderate accuracy.

Unfortunately, there is some uncertainty in the literature concerning the self-diffusion of gold. McKay[6] gives values for this coefficient which are about one hundred times those determined by Sagrubsky.[7] We have chosen to use McKay's values for two reasons. First, Sagrubsky determined diffusion coefficients for only three temperatures and the plot of log$D$ *vs* $1/T$ did not give a straight line. Second, McKay's values lie between those of the other noble metals—silver and copper. This is to be expected on the basis of the relative melting points of the three metals.

Choosing the ratio R/L to be one half (on the basis of Figs. 8 to 11), and using a value of 4.0×10^{-9} cm^2/sec for D, the expression (4) gives an effective viscosity coefficient of 0.53×10^{12} poise at 1000°C. This is to be compared with the experimental value, 1.6×10^{12} poise.

If the assumed diffusion coefficient is accurate to even an order of magnitude, the approximate agreement of expression (4) with experiment indicates that the observed rate of deformation of gold wire is not due to the migration of lattice defects between mosaic boundaries, if these lattice defects are assumed to be in equilibrium with the applied stresses at the mosaic boundaries.

One further observation tends to confirm the grain boundary mechanism. Herring[3] has suggested that the grains will be rotated axially during the first stages of deformation till the relative orientation of some of the grains becomes such that the boundary between them is no longer a source of excess lattice defects. This would lead to a strain rate decreasing with time, as observed in our experiments. If this explanation is accepted, the calculated value of expression (4) should be multiplied by a factor of 2 to 3, bringing it into better agreement with the experiment.

Herring's formula (4) predicts a decrease in viscosity coefficient with decreasing wire size. The opposite was indicated in the work of Udin,[5] *et al.*, and this constitutes the only serious objection to accepting the theory on which expression (4) is based. A definite decision on the matter could be reached by extending this type of investigation to wires of varying size or to single crystal wires.

STRUCTURES OF GOLD WIRE

It has been observed that even after heating for a few minutes at the temperatures used in the experiments described above, the gold wire had a reed-like structure

[7] A. M. Sagrubsky, Physik. Z. Sowjetunion **12**, 118 (1937).

consisting of a series of grains which ran transversely across the wire and which were about equal in length to the diameter of the wire. Most of the grain boundaries were perpendicular to the wire axis. The grain size was not appreciably changed no matter how long the wires were heated, although some other changes occurred which will be described below. Some typical examples of the microstructure of the gold wires used in these experiments are shown in Figs. 6 through 14. Figure 6 shows the structure of the wire as drawn, and Figs. 7 through 14 show the typical grain structure after stressing for the time and temperature indicated. Figures 7–9 show a phenomenon that was commonly observed in the wires which had been under low stresses; there are definite kinks at the grain boundaries. These may occur as a result of a rotation of the grains around the axis or diameter of the wire in order to achieve a low energy position for the boundary. Offsetting, or sliding of grains on their boundaries, was only observed at the higher stresses (Fig. 10).

At low stresses for long periods of time, where the specimens shortened, we could observe that the shortening was not uniform but occurred by formation of a series of waists and bulges as shown in Figs. 11 and 12. The appearance of this type of behavior may readily explain why the straining was erratic at the lower stresses. This phenomenon shown in Figs. 11 and 12 indicate that the cylindrical wires are attempting to shrink and break up into spheres in much the same manner as a column of water flowing from a spigot. In order to accentuate this behavior, a 3-in. length of smaller (0.0004-in. diameter) gold wire was laid in an alundum boat and heated at 1055°C for 120 hours. After this treatment, the original wire had broken up into 16 fragments of various lengths. The total length of the fragments was approximately 10 percent less than the length of the original single piece. Photographs of two of the fragments are shown in Figs. 13 and 14, where the waists and bulges are clearly visible.

ACKNOWLEDGMENT

The writers wish to acknowledge the help of W. E. Kingston, manager, and of D. R. Brunstetter and A. Timper, of Sylvania Metallurgical Laboratories.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 4 APRIL, 1951

Amplification of the Traveling Wave Tube*

BERNARD FRIEDMAN
Washington Square College of Arts and Sciences, Mathematics Research Group, New York University, New York
(Received September 18, 1950)

A method is presented for solving the equation which determines the modes of a helix with an electron beam [Chu and Jackson, Proc. Inst. Radio Engrs. **36**, 853 (1948)]. The essential idea is to consider the modes as a perturbation of the modes produced by the cold helix. From the solution of the equation, the dependence of the amplification factor of the tube upon the geometrical parameters and the operating conditions is explicitly determined. The solution shows also that the tube will not amplify for too large values of the dc electron beam current.

1. INTRODUCTION

IN a recent paper, Chu and Jackson,[1] have investigated the modes of an idealized traveling wave tube. One of the results of their paper is that the propagation constant γ_n associated with the nth mode satisfies a complicated transcendental equation. By making some rough approximations, the equation reduces to a cubic whose solutions can be found. However, their method is subject to the criticism that in order to perform the reduction, all the parameters[2] of the problem must be fixed. This means that when any of the parameters is changed, the whole procedure must be repeated. As a result, the effect upon the propagation constants of a change in parameters cannot be determined.

In this paper a simpler and more exact method is presented for solving the transcendental equation given by Chu and Jackson. The method yields explicitly the manner in which the propagation constants depend on the parameters so that the effect of a change of parameters can be readily calculated.

In particular a method for finding the maximum amplification of the tube, based on the above theory, is given. Graphs for the maximum amplification in terms of the ratio $I/V^{3/2}$ are given. Graphs have also been constructed for the quantity d_0 in terms of $I/V^{3/2}$, where d_0 is the difference between the wave number for the electron beam and the wave number of the cold "helix."

Finally an investigation has been made in the case

* This work was performed at Washington Square College of Arts and Sciences, New York University, and was supported in part by Contract No. AF-19(122)-1 with the U. S. Air Force through the Air Force Cambridge Research Laboratories, Air Materiel Command.

[1] L. J. Chu and J. D. Jackson, Proc. Inst. Radio Engrs. **36**, 853 (July, 1948).

[2] By parameter of the tube is meant any one of the following quantities: Geometrical characteristics—a, b radius of tube and electron beam respectively, and θ the pitch angle of the helix. Physical characteristics—I, V, dc current and voltage of the electron beam and $k^2=\omega^2\epsilon\mu$.

where ak is large. Asymptotic formulas for the bessel functions are used and the cubic equation is reduced to a simpler one. The amplification of the wave, in this case, can be explicitly written down in terms of the parameters of the tube. One result of this study is that the tube will not amplify for too large values of the dc current.

2. DERIVATION OF THE BASIC CUBIC EQUATION

In the paper by Chu and Jackson the actual helix is replaced by a lossless helical sheath of radius a and an infinitesimal thickness. The current flow along the sheath is constrained to move in a direction which makes a constant angle $90°-\theta$ with the axis of the helix. It is assumed that the electrons are confined within a cylinder of radius b concentric with the helical sheath.

Chu and Jackson define three regions: the region outside the cylinder, $r>a$, the region between the cylinder and the stream of electrons, $b<r<a$, and finally the region occupied by the electrons, $r<b$. Field components for E and H in each region are deduced from Maxwell's equations and the fields are matched on the two boundaries. If the fields are assumed to be circularly symmetric and vary with $\exp(i\omega t-\gamma z)$, the transcendental equation for the propagation constant γ is found to be[3]

$$i\frac{ka}{pa}\frac{\eta}{p}\frac{I_1(\eta b)}{I_0(\eta b)}=i\frac{ka}{pa}\left\{\frac{I_1(pb)-K_1(pb)\left[\left(\frac{ka}{pa}\right)^2\cot^2\theta I_1(pa)K_1(pa)-I_0(pa)K_0(pa)\right]}{K_0^2(pa)}\right\}\Bigg/\left\{I_0(pb)+K_0(pb)\left[\frac{\left(\frac{ka}{pa}\right)^2\cot^2\theta I_1(pa)K_1(pa)-I_0(pa)K_0(pa)}{K_0^2(pa)}\right]\right\} \tag{2.1}$$

where $I_0(x)$, $I_1(x)$, $K_0(x)$, and $K_1(x)$ are modified bessel functions, and where

$$k^2=\omega^2\epsilon\mu, \quad k=2\pi/\lambda=\omega/c, \tag{2.2}$$

$$p^2=-(\gamma^2+k^2), \tag{2.3}$$

$$\eta^2=p^2\left\{1+\frac{eI/m}{\pi b^2\epsilon v_0^3[(i\omega/v_0)-\gamma]^2}\right\}, \tag{2.4}$$

e/m = ratio of charge of electron to mass. (2.5)

v_0 = average value of the electron beam velocity, (2.6)

and

I = dc beam current. (2.7)

In order to discuss the behavior of the tube, Eq. (2.1) must be solved for p. A real root of this equation corresponds to a progressive wave while a complex root with a positive imaginary part indicates that the corresponding wave will be amplified.

A method will now be given for solving Eq. (2.1). Rewrite it in the form

$$\frac{\eta}{p}\frac{I_1(\eta b)}{I_0(\eta b)}\cdot\frac{I_0(pb)}{I_1(pb)}=\frac{1-[K_1(pb)/I_1(pb)]G(pa)}{1+[K_0(pb)/I_0(pb)]G(pa)}, \tag{2.8}$$

where

$$G(pa)=\frac{I_0(pa)}{K_0(pa)}\left[\left(\frac{ka\cot\theta}{pa}\right)^2\frac{I_1(pa)}{I_0(pa)}\frac{K_1(pa)}{K_0(pa)}-1\right]. \tag{2.9}$$

In the absence of an electron beam the modes are given by the equation

$$G(pa)=0. \tag{2.10}$$

The essential point of the present method is to consider the roots of Eq. (2.8) as being found by means of a perturbation of the roots of Eq. (2.10).

Let p_0a be the single real root of Eq. (2.10). Put

$$p_0a=x_0, \quad pa=x_0+y, \tag{2.11}$$

where y is assumed to be small compared to x_0. If the indicated division in the right-hand side of Eq. (2.8) is performed, we get a quotient of 1 and a remainder of $-G(pa)/[pbI_1(pb)I_0(pb)]$. If, furthermore, $G(pa)$ is expanded in a Taylor's series about x_0, there results

$$\begin{aligned}G(pa)=G(x_0+y)&=G(x_0)+yG'(x_0)+\cdots,\\&=yG'(x_0)+\cdots,\\&=yG'(x_0)\text{ approximately.}\end{aligned}$$

Hence

R.h.s. of Eq. (2.8)

$$=1-\frac{yG'(x_0)/[bx_0/aI_0(bx_0/a)I_1(bx_0/a)]}{1+[K_0(bx_0/a)/I_0(bx_0/a)]G'(x_0)y}. \tag{2.12}$$

For the left-hand side of the fundamental equation we proceed as follows: Expand the function $I_1(rx)/I_0(rx)$ in a Taylor's series in r about $r=1$. We find that

$$\frac{I_1(rx)}{I_0(rx)}=\frac{I_1(x)}{I_0(x)}+(r-1)x\left[\frac{I_0(x)}{I_1(x)}-\frac{I_1(x)}{I_0(x)}-\frac{1}{x}\right]\frac{I_1(x)}{I_0(x)}. \tag{2.13}$$

[3] See reference 1, pp. 853, 863, Eqs. (20), (25).

Thus

$$\frac{I_1(rx)}{I_0(rx)}\frac{I_0(x)}{I_1(x)}=1+(r-1)x\left[\frac{I_0(x)}{I_1(x)}-\frac{I_1(x)}{I_0(x)}-\frac{1}{x}\right].$$

Let $r=\eta/p$. The left-hand side of the fundamental equation can be put in the form,

$$\text{L.h.s.}=\frac{\eta}{p}+\left(\frac{\eta}{p}-1\right)\frac{\eta}{p}-\delta,$$

where

$$\delta=\left[\frac{I_0(bx_0/a)}{I_1(bx_0/a)}-\frac{I_1(bx_0/a)}{I_0(bx_0/a)}-\frac{1}{bx_0/a}\right]_a^b x_0. \quad (2.14)$$

Assuming $\eta/p=1$, Eq. (2.14) becomes

$$\text{L.h.s.}=\eta/p+[(\eta/p)-1]\delta. \quad (2.15)$$

Equating Eq. (2.12) and (2.15), our basic equation is replaced by

$$\frac{\eta}{p}+\left(\frac{\eta}{p}-1\right)\delta=1-\frac{yG'(x_0)/[bx_0/aI_0(bx_0/a)I_1(bx_0/a)]}{1+[K_0(bx_0/a)/I_0(bx_0/a)]G'(x_0)y},$$

or

$$\left(\frac{\eta}{p}-1\right)(1+\delta)=-\frac{yG'(x_0)/[bx_0/aI_0(bx_0/a)I_1(bx_0/a)]}{1+[K_0(bx_0/a)/I_0(bx_0/a)]G'(x_0)y}. \quad (2.16)$$

Since the phase velocities of the waves are much less than that of light it follows from Eq. (2.3) that $p\cong-i\gamma$ or $\gamma\cong ip$. Incorporating this result in Eq. (2.4), there results after an application of the binomial theorem that

$$\begin{aligned}\frac{\eta}{p}&=1-\frac{eI/m}{2\pi(b^2/a^2)\epsilon v_0^3(d_0-y)^2}\\&=1-\frac{1}{\pi\epsilon 4\sqrt{2}(e/m)^{\frac{1}{2}}b^2/a^2}\frac{I}{V^{\frac{1}{2}}}\cdot\frac{1}{(d_0-y)^2}\\&=1-\frac{W}{2b^2/a^2}\frac{1}{(d_0-y)^2},\end{aligned} \quad (2.17)$$

where we have used the abbreviations

$$d_0=\omega a/v_0-x_0 \quad (2.18)$$

$$W=\frac{e/m}{\pi\epsilon}\frac{I}{v_0^3}=\frac{1}{\pi\epsilon 2\sqrt{2}(e/m)^{\frac{1}{2}}}\frac{I}{V^{\frac{1}{2}}} \quad (2.19)$$

since $1/2mv_0^2=eV$. Combining Eqs. (2.16) and (2.17) we get

$$y(y-d_0)^2-\zeta(1+\xi y)=0 \quad (2.20)$$

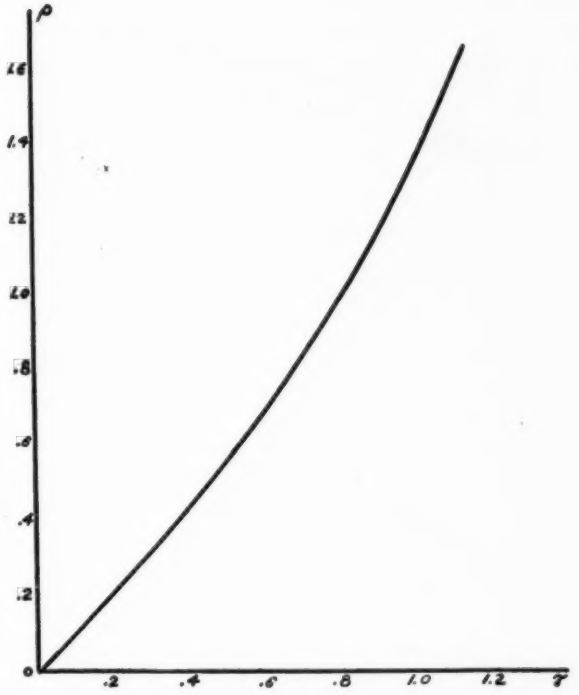


FIG. 1. Graph of $\rho^3=\tau^2(1+\rho\tau)$.

where

$$\zeta=\frac{x_0I_0(bx_0/a)I_1(bx_0/a)(1+\delta)W}{2b/aG'(x_0)} \quad (2.21)$$

$$\xi=\frac{K_0(bx_0/a)}{I_0(bx_0/a)}G'(x_0). \quad (2.22)$$

Equation (2.20) is the desired cubic equation, ζ and ξ are known quantities depending on the geometry of the tube, and the voltage and current of the beam. In the case where the electron beam travels at such a speed that it is in exact step with the wave traveling along the helix, we have $d_0=0$. Thus d_0 may be considered a measure of how far from resonance the tube is.

The cubic equation (2.20) has real coefficients. It has three real roots, or one real root and a pair of complex roots. The case of three real roots (γ pure imaginary) corresponds to modes which are neither attenuated nor amplified. In the other case, however, the pair of complex conjugate roots gives rise on the one hand to an attenuated wave, on the other hand to an amplified wave. The amount of attenuation and amplification depends on the imaginary part of the complex roots of Eq. (2.20). We shall investigate this question in the next section.

3. MAXIMUM AMPLIFICATION OF THE TUBE

The question of what values of the parameter d_0 give maximum amplification will now be taken up. It will be shown that maximum amplification occurs for a value of $d_0<0$ so that the velocity of the electron beam should

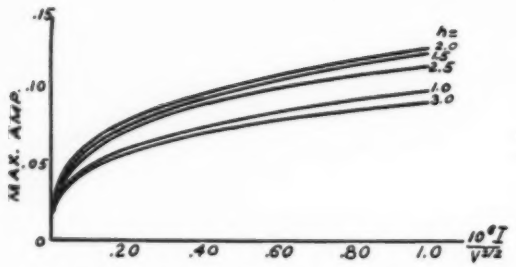


FIG. 2a. Maximum amplification as a function of $10^6 I/V^{3/2}$. $b/a=0.6$, $h=ak\cot\theta$.

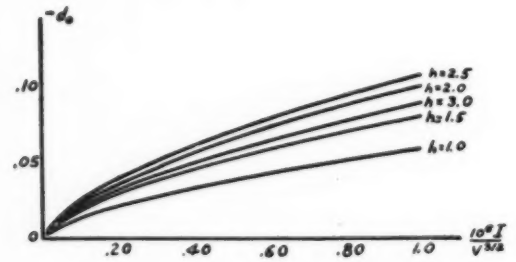


FIG. 2b. $-d_0$ as a function of $10^6 I/V^{3/2}$. $b/a=0.6$, $d_0=(a\omega/v_0)-x_0$, $h=ak\cot\theta$.

be slightly greater than that of the wave down the tube. But the maximum is so broad that $d_0=0$ will give amplification close enough to maximum.

We start with our basic equation

$$y(y-d_0)^2-\zeta(1+\xi y)=0 \tag{2.20}$$

and pose the problem of determining the value of d_0 which will give maximum amplification.

Let

$$y=z\zeta^{\frac{1}{3}},\quad \tau=d_0/\zeta^{\frac{1}{3}},\quad \rho=\xi\zeta^{\frac{1}{3}} \tag{3.1}$$

then Eq. (2.20) becomes

$$z(z-\tau)^2-(1+\rho z)=0, \tag{3.2}$$

and we ask what value of τ will yield the maximum imaginary part for a complex root of Eq. (3.2).

Let $z=\alpha+i\beta$, then as we will show in the Appendix τ is given by

$$\tau^3(1+\rho\tau)=\rho^3, \tag{3.3}$$

and maximum β by

$$\beta^2=1/4\left(\frac{3\rho}{\tau}+\tau\right)\left(\frac{\rho}{\tau}-\tau\right). \tag{3.4}$$

Equations (3.3) and (3.4) provide the solution of our problem. A graph of τ as a function of ρ is given in Fig. 1. In terms of the original equation, Eq. (3.4), becomes

$$\beta^2\zeta^{2/3}=1/4\left(\frac{3\xi\zeta}{d_0}+d_0\right)\left(\frac{\xi\zeta}{d_0}-d_0\right),$$

where d_0 is determined by $d_0=\tau\zeta^{\frac{1}{3}}$ [see Eq. (3.1)].

Figures 2 and 3 represent graphs of d_0 and maximum amplification as functions of the ratio $10^6 I/V^{\frac{3}{2}}$.

4. DISCUSSION FOR LARGE VALUES OF pa

Further discussion of the problem is possible if it is assumed that pa is large, say, about 3 or greater. The Eq. (2.1) again reduces to a cubic equation but since the bessel functions may, in this case, be replaced by their asymptotic values, the coefficients ξ and ζ of Eq. (2.20) are considerably simplified.

First we note that Eq. (2.9) becomes

$$(pa/ka)^2\tan^2\theta=1 \tag{4.1}$$

so that

$$x_0=ka\cot\theta. \tag{4.2}$$

Assume that $d_0=0$, that is that the electron beam is in step with the wave. Equation (2.20) reduces to

$$y^3-\zeta\xi y-\zeta=0. \tag{4.3}$$

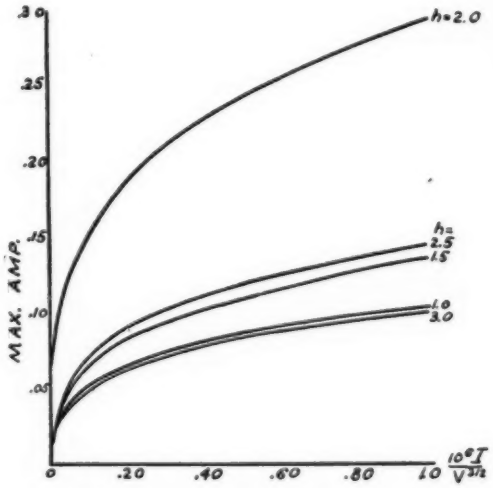


FIG. 3a. Maximum amplification as a function of $10^6 I/V^{3/2}$. $b/a=0.8$.

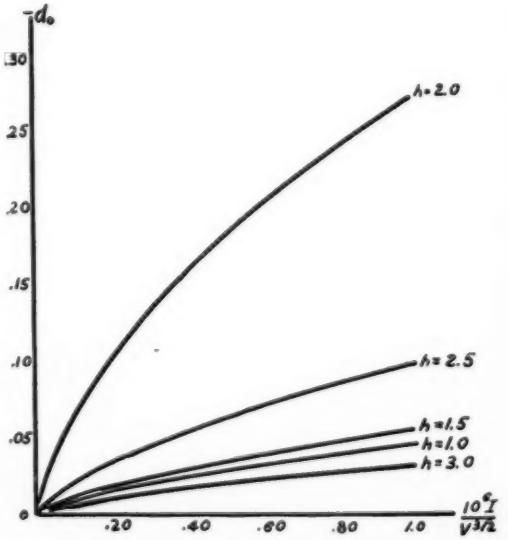


FIG. 3b. $-d_0$ as a function of $10^6 I/V^{3/2}$. $b/a=0.8$.

If use is made of the asymptotic formula

$$I_\nu(x)\cong\frac{1}{(2\pi x)^{\frac{1}{2}}}e^x,\quad K_\nu(x)\cong\left(\frac{\pi}{2x}\right)^{\frac{1}{2}}e^{-x},\quad \nu=0,1, \tag{4.4}$$

in the coefficients ζ and ξ, Eq. (4.3) becomes

$$Ay^3-LAy+L=0 \tag{4.5}$$

where

$$L=\left(\frac{a}{b}\right)^2\cdot\frac{\sqrt{2}}{16\pi\epsilon(e/m)^{\frac{1}{2}}}\frac{I}{V^{3/2}},$$

$$A=\frac{2}{x_0}\exp\left[-2x_0\left(\frac{b}{a}-1\right)\right]. \tag{4.6}$$

Let $y=(L/A)^{\frac{1}{2}}u$, then Eq. (4.5) becomes

$$u^3-Bu+1=0,\quad B=(A^2L)^{\frac{1}{2}}. \tag{4.7}$$

Equation (4.7) will have imaginary roots if $B<3/4^{\frac{1}{2}}$. The quantity B, however, contains the term $I/V^{\frac{1}{2}}$. Thus if all other quantities are kept constant, we see that no amplification occurs if the above ratio becomes too large.

Another consequence of Eq. (4.7) is as follows: Suppose $B\ll1$. This implies that for the real root u is approximately -1. A better approximation is given by

$$u=-(1+B)^{\frac{1}{2}}=-1-B/3+\cdots, \tag{4.8}$$
$$=-1-(B/3).$$

But if the real root of Eq. (4.7) is given by Eq. (4.8), the quadratic equation yielding the complex roots is

$$u^2-u[1+(B/3)]+[1-(B/3)]=0 \tag{4.9}$$

or

$$u=\frac{1+B/3\pm(-3+2B)^{\frac{1}{2}}}{2}. \tag{4.10}$$

Hence the imaginary part is given by

$$Imu=\pm\frac{(3-2B)^{\frac{1}{2}}}{2}=\pm\frac{\sqrt{3}}{2}(1-\tfrac{1}{3}B).$$

An example of the use of the results of this section will be given. Let $x_0=ka\cot\theta=2.4$; $b/a=0.8$, $I=25ma$, $V=2500$. Then $L=2.35\cdot10^{-3}$, $A=2.18\cdot(L/A)^{\frac{1}{2}}=1.024\cdot10^{-1}$, $B=(A^2L)^{\frac{1}{2}}=2.2\cdot10^{-1}$ and $1/2(3-2B)^{\frac{1}{2}}=0.80$. Hence $y=(L/A)^{\frac{1}{2}}u$ gives for the imaginary part of the root of Eq. (4.5)

$$Imy=0.0820.$$

This is in agreement with the graph in Fig. 3 for the values of the $b/a=0.8$ and $10^6I/V^{\frac{1}{2}}=0.2$. Since d_0 is small, the amplification could be computed by the results of this section rather than those of Sec. 3.

Referring back to the graphs in Figs. 2 and 3, it is seen that the curves for maximum amplification reach a maximum for ($b/a=0.8$, $h=2.0$) and ($b/a=0.6$, $h=2.5$). An explanation of this is afforded by the work of the present section.

It is readily verified that if b/a is fixed and if $0<b/a<1$, the function A of Eq. (4.6) has a minimum for $x_0=1/[2(1-b/a)]$. The quantity L of Eq. (4.6) is independent of x_0. Thus it is possible for the roots of Eq. (4.5) to be the same for two different values of x_0. In the work of Sec. 3, the quantity h is close to x_0.

This remark may have practical value for experiment since it indicates that the pitch angle should be so chosen that $ka\cot\theta=2.0$ for $b/a=0.8$ and $ka\cot\theta=2.5$ for $b/a=0.6$.

5. APPENDIX

We want to determine the value of τ in the cubic equation

$$z(z-\tau)^2-(1+\rho z)=0\quad(\tau,\ \rho\text{ real}) \tag{5.1}$$

so that the imaginary part of the complex root is a maximum. Let $z=\alpha+i\beta$, then the imaginary part of Eq. (5.1) may be written

$$\beta^2=3\alpha^2-4\alpha\tau+\tau^2-\rho. \tag{5.2}$$

Let z_0 be the real root of Eq. (5.1), then $z_0+2\alpha=2\tau$ or $\alpha=\tau-z_0/2$. Differentiating Eq. (5.2) with respect to τ yields $d\beta^2/d\tau=6\alpha d\alpha/d\tau-4\tau d\alpha/d\tau-4\alpha+2\tau=0$ for a maximum, or

$$\frac{d\alpha}{d\tau}=\frac{2\alpha-\tau}{3\alpha-2\tau}=\frac{\tau-z_0}{\tau-\frac{3}{2}z_0}.$$

But

$$\frac{d\alpha}{d\tau}=1-\frac{1}{2}\frac{dz_0}{d\tau},$$

and

$$3z_0^2\frac{dz_0}{d\tau}-4\tau z_0\frac{dz_0}{d\tau}+(\tau^2-\rho)\frac{dz_0}{d\tau}-2z_0^2+2\tau z_0=0$$

so that

$$\frac{dz_0}{d\tau}=\frac{2z_0^2-2\tau z_0}{3z_0^2-4\tau z_0+\tau^2-\rho}=2\left(1-\frac{d\alpha}{d\tau}\right)=-\frac{z_0}{\tau-\frac{3}{2}z_0}. \tag{5.3}$$

Equation (5.3) reduces to

$$2\tau z_0=2\tau^2+2\rho$$

or

$$z_0=\tau+(\rho/\tau). \tag{5.4}$$

Using this value of z_0 given by Eq. (5.4) in Eq. (5.1) gives us an equation which reduces to

$$\tau^3(1+\rho\tau)=\rho^3. \tag{5.5}$$

Equation (5.5) is the solution of our problem. Now use Eq. (5.5) in Eq. (5.2) to get

$$\beta^2=\frac{1}{4}\left(\frac{3\rho}{\tau}+\tau\right)\left(\frac{\rho}{\tau}-\tau\right). \tag{5.6}$$

Equation (5.6) provides the maximum imaginary part for a given τ.

ACKNOWLEDGMENT

The author is deeply indebted to Dr. Henry Malin for his assistance in the preparation of this paper.

Interfacial Free Energy of Coherent Twin Boundaries in Copper

R. L. FULLMAN*
General Electric Research Laboratory, Schenectady, New York
(Received October 26, 1950)

The ratio of the interfacial free energy of coherent twin boundaries to the average grain boundary free energy in copper has been measured by means of the dihedral angles formed at the intersections of twin boundaries and grain boundaries with each other and with a copper/lead vapor interface. The mean value found for the ratio is 0.035 ± 0.006.

TWO crystals are said to have a twinned orientation relationship with respect to each other when the orientation of one is a mirror image of that of the other, across a plane called the twinning plane. Either crystal may be termed a twin of the other. The boundary joining the crystals is called the "twin boundary," or "composition plane."

Twinned crystals have been observed in most metals, either in the deformed or annealed condition, or both. It has been common to refer to the twins formed directly as a result of deformation as "deformation twins" or "mechanical twins," and to call those found in annealed metals "annealing twins." The observations of annealing twins in metals have been reviewed extensively by Carpenter and Tamura[1] and by Mathewson.[2]

The boundary between twinned crystals frequently coincides with the twinning plane that relates their orientations. Since neighboring atoms can match up perfectly across a twin boundary that coincides with the {111} twinning plane of face-centered cubic crystals, it is suggested that a boundary with this orientation be called a "coherent twin boundary." Twin boundaries that do not coincide with the twinning plane do not permit exact matching of neighboring atom positions across the boundary, and such boundaries may be termed "noncoherent twin boundaries."

It is frequently observed that the dihedral angles between phase and grain interfaces, at the lines and points where they meet, are determined by a mechanical balance between the interfacial tensions. This mechanical balance is equivalent to minimizing the interfacial free energy of the system, and is attained in solids as well as in liquids if equilibrium is established in the region under consideration. Smith[3] has shown that many features of microstructures may be interpreted as the result of equilibrium of interfacial tensions, and has determined the relative values of many grain and phase boundary free energies. Dunn and Lionetti[4] have investigated the dependence of grain boundary free energy on relative orientation of the grains, assuming the influence of boundary orientation to be negligible. The variation of grain boundary free energy with relative grain orientation and with boundary orientation has been treated in theoretical papers by Herring[5] and by Read and Shockley.[6] Dunn, Daniels, and Bolton[7] have investigated the interfacial free energy of twin boundaries in silicon iron, using an approximate method based on Herring's analysis.

The interfacial free energy of a boundary between a face-centered cubic crystal and its annealing twin depends strongly on the boundary orientation. The perfect fit of nearest neighbor atoms across a boundary that coincides with the {111} twinning plane suggests that such a boundary would have a much lower interfacial free energy than a boundary of different orientation. This notion is confirmed by the observation that annealing twins are bounded predominantly by plane surfaces coinciding with the twinning plane. Thermal etching experiments on silver carried out by Shuttleworth, King and Chalmers[8,9] show that the grooves formed at twin boundaries are much fainter at boundaries coinciding with the twinning plane than at twin boundaries of other orientation. The grooves formed on thermal etching are the result of equilibration of the interfacial tensions at the junctions of solid-solid interfaces with the free surface. Hence, solid-solid interfaces with relatively small interfacial tension are revealed by comparatively shallow, or faint, grooves. Read and Shockley[6] suggest that the variation of twin boundary free energy with boundary orientation should exhibit a cusp at the orientation of the twinning plane.

MEASUREMENT OF THE INTERFACIAL FREE ENERGY OF COHERENT TWIN BOUNDARIES BY MEANS OF THE DIHEDRAL ANGLES FORMED AT THEIR INTERSECTIONS WITH GRAIN BOUNDARIES

At the junction of a coherent twin boundary and a grain boundary, the dihedral angle opposite the twin

* This paper is based upon a part of a dissertation submitted by R. L. Fullman to the faculty of the Engineering School of Yale University in partial fulfillment of the requirements for the degree of Doctor of Engineering, June, 1950.

[1] H. C. H. Carpenter and S. Tamura, Proc. Roy. Soc. (London) **113A**, 161 (1926).

[2] C. H. Mathewson, Proc. Inst. Metals Div., Am. Inst. Mining Met. Engrs. **1928**, 7 (1928).

[3] C. S. Smith, Trans. Am. Inst. Mining Met. Engrs. **175**, 15 (1948).

[4] C. G. Dunn and F. Lionetti, Trans. Am. Inst. Mining Met. Engrs. **185**, 125 (1949).

[5] C. Herring, "Surface tension as a motivation for sintering" (to be published).

[6] W. T. Read and W. Shockley, "Dislocation models of crystal grain boundaries" (to be published).

[7] Dunn, Daniels, and Bolton, Trans. Am. Inst. Mining Met. Engrs. **188**, 368 (1950).

[8] Shuttleworth, King, and Chalmers, Nature **158**, 482 (1946).

[9] Chalmers, King, and Shuttleworth, Proc. Roy. Soc. (London) **193A**, 465 (1948).

boundary may be either greater or less than 180°, as shown in Fig. 1. The configuration at point A is not consistent with a mechanical balance of interfacial tensions that are independent of interface orientation. In order to explain the occurrence of interfacial dihedral angles greater than 180°, Smith[3] proposed that the interfacial tension of a twin boundary is small in comparison with the variation in interfacial tension of grain boundaries with change in boundary orientation, and that each grain boundary is oriented so as approximately to minimize its own interfacial free energy. At the junction of an annealing twin boundary with a grain boundary, the orientation relationship across the grain boundary changes discontinuously, accompanied by an abrupt change in the boundary orientation.

The occurrence of interfacial dihedral angles greater than 180° can be explained on another basis. In the absence of detailed knowledge concerning the manner in which grain boundary free energy depends on the relative orientation of the grains and on the boundary orientation, it is possible to rationalize the observed configurations by assuming that no significant contribution arises from the variation of grain boundary free energy with boundary orientation.[10] The interpretation is based on Burke's observation[11] that coherent twin boundaries do not migrate at a perceptible rate during annealing. Since the orientations of a grain and its twin completely determine the orientation of the twin boundary, it follows that its intersection with a grain boundary cannot move readily except in a direction parallel to the twin boundary plane. Thus, during annealing, the configuration of interfaces reaches equilibrium only with respect to interfacial tension components parallel to the twin boundary. Referring to Fig. 1 let σ_{12}, σ_t, and σ_{t2} represent the interfacial tensions of the boundaries between grains 1 and 2, between grain 1 and its twin, and between grain 2 and the twin of grain 1. The configurations at points A and B may be obtained if σ_{12} is slightly larger than σ_{t2}, and both σ_{12} and σ_{t2} are much larger than σ_t. A mechanical analog of the surface tensions may be constructed as shown in Fig. 2, with the grain boundaries represented by strings and the twin boundaries represented by rigid rods that can move only in their longitudinal direction.

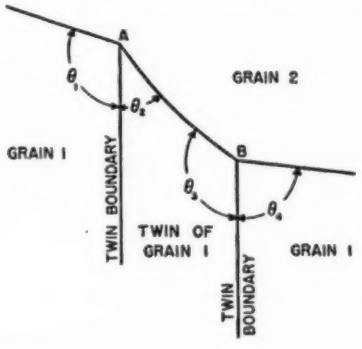


FIG. 1. Intersection of an annealing twin and a grain boundary. The plane of the drawing is perpendicular to the interfaces.

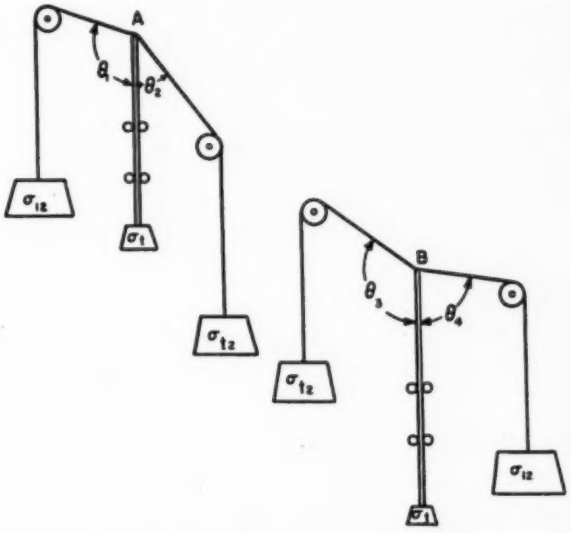


FIG. 2. Mechanical analog of surface tensions acting at the intersection of an annealing twin and a grain boundary.

The condition that, at equilibrium, surface tension components parallel to the twin boundaries are balanced, may be written as follows:

$$\begin{aligned} &\text{Point } A, \quad \sigma_{12}\cos\theta_1+\sigma_{t2}\cos\theta_2+\sigma_t=0.\\ &\text{Point } B, \quad \sigma_{12}\cos\theta_4+\sigma_{t2}\cos\theta_3+\sigma_t=0. \end{aligned} \tag{1}$$

This relationship is identical to that derived by Herring[5] on the basis of minimization of interfacial free energy with respect to displacements of the junction point along the twin boundary for the condition that σ_{12} and σ_{t2} are independent of boundary orientation. Dunn, Daniels, and Bolton[7] utilized an equation of this form, along with the assumption that $\sigma_{12}=\sigma_{t2}$, to estimate the ratio σ_t/σ_{12} from the angles formed at the intersection of a twin boundary and a grain boundary in silicon iron.

Since only two grain boundary interfacial tensions appear in Eq. (1), the ratio between the interfacial tension of the twin boundary and that of either grain boundary may be calculated from measurements of all four dihedral angles. These ratios are:

$$\frac{\sigma_t}{\sigma_{12}}=\frac{\cos\theta_2\cos\theta_4-\cos\theta_1\cos\theta_3}{\cos\theta_3-\cos\theta_2}, \tag{2}$$

$$\frac{\sigma_t}{\sigma_{t2}}=\frac{\cos\theta_2\cos\theta_4-\cos\theta_1\cos\theta_3}{\cos\theta_1-\cos\theta_4}. \tag{3}$$

EXPERIMENTAL METHOD

When a solid in the form of thin sheet is annealed, interfaces that extend completely through the sheet thickness tend to become perpendicular to the sheet surface if the interfacial free energies of the boundaries involved are nearly independent of their orientations. Thus, the dihedral angles between the interfaces may be

[10] J. P. Nielsen, discussion of reference 13.
[11] J. E. Burke, Trans. Am. Inst. Mining Met. Engrs. 188, 1324 (1950).

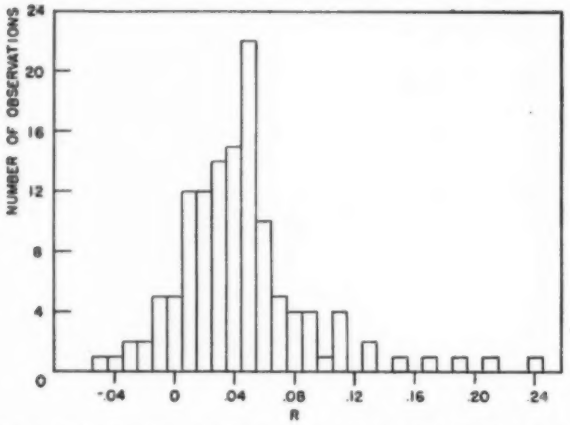


FIG. 3. Distribution of values of R. Mean 0.045. Probable error of the mean 0.0030.

determined by measuring the angles between their traces on the sheet surface. This technique has been used by Dunn.[4,12] A pronounced texture in the sheet permitted use of the same technique in the present investigation, despite the strong dependence of the twin boundary free energy on the boundary orientation.

If a grain has a {110} plane parallel to the sheet surface, two of the {111} twinning planes are perpendicular to the sheet surface and the other two make an angle of about 35° with it. If a twin has one of the {111} planes that is perpendicular to the surface as its twin boundary, the twin also has a {110} plane parallel to the sheet surface, so that any possible variation of solid-vapor surface tension with surface orientation cannot influence the interfacial configurations. If only twins with boundaries nearly perpendicular to the sheet surface are used, the angles between traces of the twin boundaries and grain boundaries on the sheet surface (θ_i') are nearly equal to the corresponding dihedral angles (θ_i).

The present investigation was carried out on compression-rolled† sheet with a pronounced {110} texture. The dihedral angle ψ between the twin boundary and the sheet surface was greater than 77° for all twins used. The interfacial dihedral angles were computed from the relationship

$$\cos\theta_i \simeq \sin\psi \cos\theta_i'. \quad (4)$$

A specimen of OFHC copper bus-bar 0.200 in. in thickness was annealed at 500°C for one hour and compression rolled to 0.020 in. with two intermediate anneals at 500°C. The specimen was then annealed at 950°C for 40 hours. The resulting structure consisted of grains extending completely through the sheet, with many annealing twins. Most of the grain boundaries were approximately perpendicular to the plane of the sheet. The specimen was electrolytically polished and etched. The angles θ_1', θ_2', θ_3', θ_4' were measured at 27 twin-grain boundary intersections at a magnification of 770×, using the rotating table of a metallograph. Each angle was measured three times by each of three observers. The specimen was then strained to produce slip lines on the surface. The orientations of the twinned grains and the values of ψ were determined from the angles between the slip lines, and the proximity of ψ to 90° was checked by the appearance under polarized light after etching with concentrated nitric acid. Values of θ_1, θ_2, θ_3, θ_4 were calculated from the corresponding values of θ', by means of Eq. (4).

Values of the ratios σ_i/σ_{12} and σ_i/σ_{i2} were calculated by means of Eqs. (2) and (3). Figure 3 is a block diagram of the frequency of occurrence of various values of these ratios (R). The mean value of R is 0.045 ± 0.0030.

MEASUREMENT OF THE INTERFACIAL FREE ENERGY OF COHERENT TWIN BOUNDARIES BY MEANS OF THE DIHDERAL ANGLES FORMED AT THEIR INTERSECTIONS WITH A COPPER/LEAD VAPOR INTERFACE

Markings appearing on a polished metal surface after annealing in vacuum, and indicating the location of grain boundaries, were reported 40 years ago;[13] and Chalmers, King, and Shuttleworth[8,9] recently demonstrated that the markings developed on silver as a result of thermal etching are grooves in the surface. They proposed that the grooves result from a balance of the surface tension of the solid-solid interface revealed with that of the free surface. It was observed that the markings revealing coherent twin boundaries are much fainter than those at either noncoherent twin boundaries or grain boundaries.

Experiments conducted by Sears[14] suggested that the surface tension of solid copper might be greatly decreased in the presence of lead vapor. It seemed possible that thermal etching in the presence of lead vapor would develop grooves with sufficiently small dihedral angles to permit measurement, in cross section, of the angles formed where coherent twin boundaries, as well as grain boundaries, meet the surface. Measurements of grain boundary groove angles in copper[15] annealed in unsaturated lead vapor and in other atmospheres have been reported since completion of the present investigation, and confirm the decrease in the surface tension of copper in the presence of lead vapor.

A specimen of OFHC copper bus-bar 0.200 in. in thickness was annealed at 500°C for one hour and compression rolled to 0.020 in. with two intermediate anneals at 500°C. After a final anneal for 1 hour at 650°C the sheet was mechanically polished. It was supported above a bath of a copper-lead alloy, containing an excess of copper, in the manner shown in Fig. 4. The

[12] C. G. Dunn, Trans. Am. Inst. Mining Met. Engrs. **185**, 72 (1949).

† In compression rolling the rolling direction is changed at each pass of the sheet through the rolls.

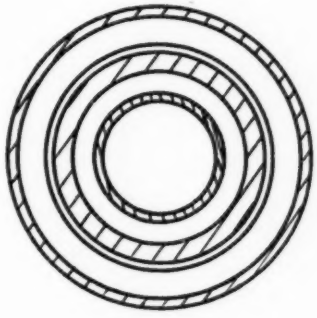
[13] W. Rosenhain and J. C. W. Humphrey, Proc. Roy. Soc. (London) **83A**, 200 (1910).

[14] G. W. Sears, J. Appl. Phys. **21**, 721 (1950).

[15] G. L. J. Bailey and H. C. Watkins, Proc. Phys. Soc. (London) **63B**, 350 (1950).

container is a quartz tube which extends upward, from the portion shown, to a graded seal with a Pyrex tube. The tube was evacuated and flushed out with hydrogen and finally sealed off with a partial atmosphere of hydrogen present, in order to restrict the evaporation of lead. The quartz lid placed above the copper specimen served to isolate a small region of approximately uniform temperature, and to permit the equilibrium vapor pressure of the lead to be maintained in this region.

Figure 5 shows the structure revealed on the surface of a polished copper specimen annealed in this manner for 100 hours at 945°C. Positions where grain boundaries had formed grooves, before further boundary migration, are revealed by grooves that are slightly rounded, and



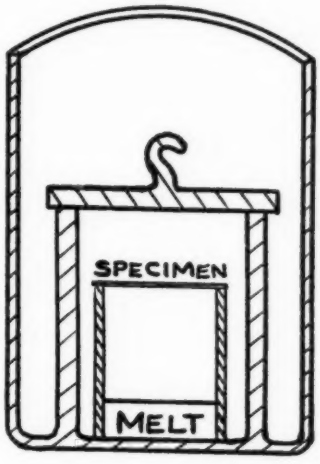


FIG. 4. Experimental arrangement for annealing polished copper sheet in an atmosphere of lead vapor.

present a less sharp appearance than the grooves where boundaries are now located. Similar rounded grooves may be observed at previous positions of noncoherent twin boundaries which have migrated in such a direction as to decrease the area of coherent twin boundary present.

A preliminary estimate was made of the groove angle to be expected at the intersection of a twin boundary with the free surface, based on Sears' estimate[14] of the surface tension of copper in the presence of lead vapor and on the previous investigation of the twin boundary to grain boundary surface tension ratio. Since at best the grooves formed would be expected to be quite shallow, it was decided that an oblique sectioning technique should be used for making the measurements. An

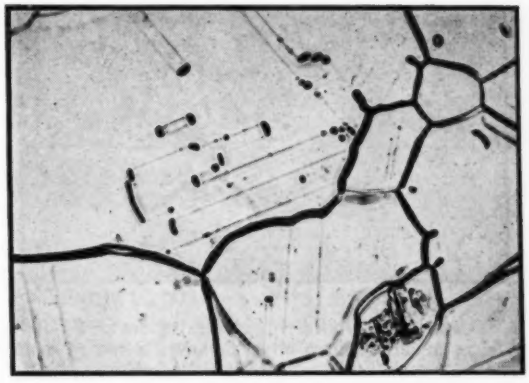


FIG. 5. Surface of compression rolled copper sheet after annealing for 100 hours at 945°C in an atmosphere of lead vapor. Thermal etch. 25×.

analysis of the relative errors introduced by uncertainty in measuring the groove angles and uncertainty in the angle of the cross section revealed that there is an optimum sectioning angle for most accurate measurement of a particular set of grooves. This analysis, along with other aspects of the analysis of data measured on oblique cross sections, is presented in the Appendix. For grooves with the expected dihedral angle for coherent twin boundaries meeting the free surface, it was determined that the optimum angle for the cross section was at an angle of 10° from the initial surface.

The specimen was copper plated in order to prevent rounding off of the edge during subsequent polishing. The specimen was then mounted in plastic, and polished sections were prepared with the aid of a polishing jig designed to provide an angle of 10° between the original surface and the polished sections. Subsequent measurements indicated that this angle was actually 8.79° ±0.06°. Figure 6 shows the appearance of grooves formed at the intersections of a grain boundary and coherent twin boundaries with the specimen surface. Figure 7 shows the type of rounded grooves remaining on the surface after migration of a grain boundary.

Measurements of the groove angles were made by

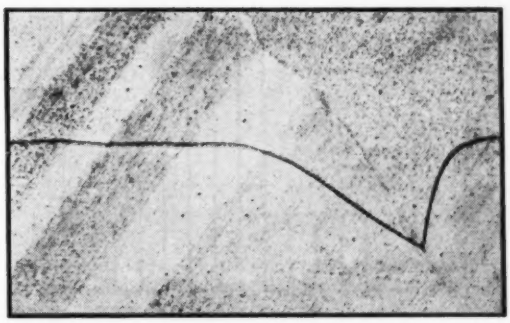


FIG. 6. Compression-rolled copper sheet annealed for 100 hours at 945°C in lead vapor. Intersection of a grain boundary (right) and coherent twin boundaries (left) with the copper/lead (vapor) interface. Section of observation is at an angle of 8.79° to the specimen surface. Ammonium peroxide etch. 500×.

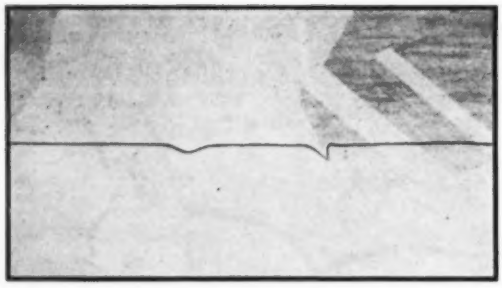


FIG. 7. Compression-rolled copper sheet annealed for 100 hours at 945°C in an atmosphere of lead vapor. Rounded depression (left) in the copper/lead (vapor) interface indicates the position of a groove formed where a grain boundary met the interface some time before completion of the anneal. The sharp groove at the right indicates the present position of an intersection of a grain boundary with the copper/lead (vapor) interface. Section of observation is at an angle of 8.79° to the specimen surface. Ammonium peroxide etch. 125×.

means of the rotating table of the metallograph, at a magnification of approximately 1100×. The angular position of the metallograph table was measured, when the cross hair of the eyepiece was parallel to each side of the groove, to the grain or twin boundary trace, and to the line of intersection between the plane of section and the plane of the original surface. These measurements determine the values of the angle ϕ, between the traces of the groove sides, and the angle β, between the boundary trace and the intersection of the plane of observation with the plane of the specimen surface. With the preferred orientation present in the copper sheet used, nearly all the twin boundaries, as well as the grain boundaries, are approximately perpendicular to the specimen surface. For a boundary that is perpendicular to the surface the dihedral angle between the groove sides is completely defined by the values of ϕ and β, if the surface tension of the copper/lead vapor interface is independent of boundary orientation. The dihedral angles, θ, formed at the grooves were calculated by means of Eq. (9A) of the Appendix. Figure 8 is a block diagram of the distribution of dihedral angles, θ_G, found at grooves where grain boundaries meet the copper/lead vapor interface. Figure 9 is a similar diagram for the angles θ_t formed at twin boundary grooves. The mean values found were $\theta_G = 142.0° \pm 1.4°$ and $\theta_t = 179.04° \pm 0.07°$.

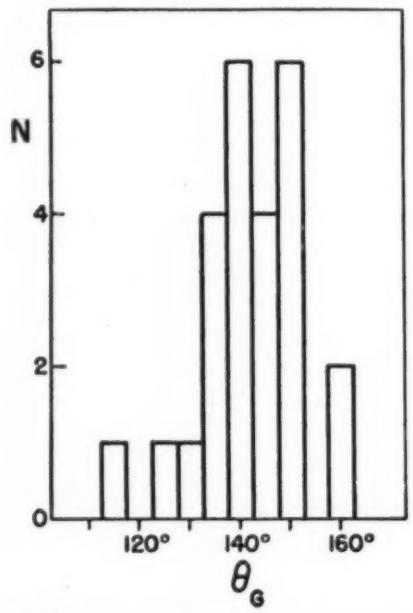


FIG. 8. Distribution of dihedral angles of grooves formed at intersections of grain boundaries with a copper/lead (vapor) interface after annealing for 100 hours at 945°C.

If the interfacial free energy of the copper/lead vapor interface, $\sigma_{Cu-Pb(v)}$, is independent of interface orientation, the relative interfacial free energies are given by

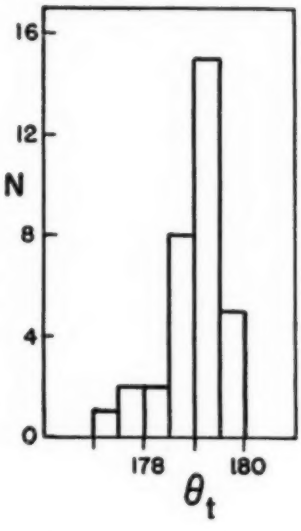


FIG. 9. Distribution of dihedral angles of grooves at intersections of coherent twin boundaries with a copper/lead (vapor) interface after annealing for 100 hours at 945°C.

the following relationships:

$$\frac{\sigma_G}{\sigma_{Cu-Pb(v)}} = 2\cos\frac{\theta_G}{2}. \tag{5}$$

$$\frac{\sigma_t}{\sigma_{Cu-Pb(v)}} = 2\cos\frac{\theta_t}{2}. \tag{6}$$

$$\frac{\sigma_t}{\sigma_G} = \frac{\cos\theta_t/2}{\cos\theta_G/2}. \tag{7}$$

Substitution of the mean values of the dihedral angles in these equations leads to the following mean values for the relative interfacial free energies:

$$\frac{\sigma_G}{\sigma_{Cu-Pb(v)}} = 0.651 \pm 0.023.$$

$$\frac{\sigma_t}{\sigma_{Cu-Pb(v)}} = 0.0168 \pm 0.0013.$$

$$\frac{\sigma_t}{\sigma_G} = 0.0257 \pm 0.0022.$$

Some systematic error has led to a discrepancy between the values of σ_t/σ_G found by observations of twin boundary-grain boundary intersections and by observations of thermally etched grooves. In the absence of any knowledge concerning the origin of the systematic error, it appears reasonable to weight the two measurements equally. The mean value is then $\sigma_t/\sigma_G=0.035\pm0.006$.

Data collected from several sources[16] indicate that the average grain boundary free energy in copper is about 600 ergs cm^{-2}. Accepting this value, the interfacial free energy of coherent twin boundaries in copper is about 21 ergs cm^{-2}, according to the combined results of the two investigations. Heidenreich and Shockley[17] have estimated on theoretical grounds that the coherent twin boundary free energy is of the order of 20 ergs cm^{-2}.

To a first approximation, a coherent twin boundary in a face-centered cubic structure may be considered as three close-packed layers of a hexagonal close-packed structure. On the basis of this concept, the free energy difference between the face-centered cubic and hexagonal close-packed structures of copper, corresponding to an interfacial free energy of 21 ergs cm^{-2}, is about 58 cal/g atom.

CONCLUSIONS

The ratio of the interfacial free energy of coherent twin boundaries to the average grain boundary free energy in copper has been determined by means of measurements of the dihedral angles formed at the intersections of twin boundaries and grain boundaries with each other and with a copper/lead vapor interface. The mean value found for the ratio is 0.035 ± 0.006.

ACKNOWLEDGMENT

The author is indebted to many members of the Metallurgy Division of the General Electric Research Laboratory for assistance, and in particular to J. C. Fisher and G. W. Sears. The investigation was supported by the U. S. Atomic Energy Commission.

APPENDIX

Analysis of Data Obtained by Oblique Sectioning

Selection of the Optimum Sectioning Angle

Figure 10 represents the balance of surface tensions of a solid-solid interface, σ_1, and of the interfaces between the solid and a second phase, σ_2, σ_3. For simplicity it will be assumed that the free energy of the interface between the solid and the second phase is independent of the orientation of the solid, so that $\sigma_2=\sigma_3$. Then angles θ_2 and θ_3 are equal and the ratio of the interfacial free energies is defined by θ_1 alone.

$$R\equiv\sigma_1/\sigma_2=2\cos\tfrac{1}{2}\theta_1. \tag{1A}$$

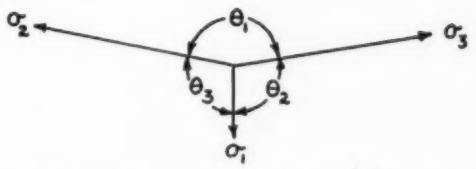


FIG. 10. Diagram illustrating the interfacial free energies of three interfaces and the dihedral angles formed at their intersection.

The precision with which the ratio of two interfacial free energies can be measured, for a given precision in measuring the dihedral angles, increases as the value of the ratio increases. For the simple situation described above, the relative error in R, per unit error in measuring θ_1, is

$$\frac{dR/R}{d\theta_1}=-\frac{\tan\frac{1}{2}\theta_1}{2}=-\frac{(4-R^2)^{\frac{1}{2}}}{2R}. \tag{2A}$$

Except for large values of R (θ_1 near zero), the relative error in measuring R per unit error in measuring θ_1 (expressed in radians^{-1}) is nearly equal to $1/R$. Figure 11 shows the relationship expressed in various units. When the ratio of the interfacial free energies to be compared is small (values of θ_1 near 180°), one must measure the dihedral angle very precisely to obtain reasonable accuracy.

The precision with which dihedral angles near 180° may be measured can be increased by making the observations on traces in a plane that is inclined at a small angle to the surface in which the grooves lie. Consider the specimen shown in Fig. 12, in which a section of observation, S_2, makes an angle γ with the original surface, S_1. For simplicity it is assumed that the groove in the surface S_1 is perpendicular to the line of intersection between S_1 and S_2, and that the bisector of the groove dihedral angle is perpendicular to S_1. Then the dihedral angle (θ_1) between the sides of the groove is related to the angle (ϕ_1) between the traces of the groove sides on the section S_2 by the relationship

$$\tan\frac{\theta_1}{2}=\frac{\tan\frac{1}{2}\phi_1}{\sin\gamma}. \tag{3A}$$

The error $d\theta_1$ in θ_1, corresponding to small errors $d\phi_1$ and $d\gamma$ in the measurement of ϕ_1 and γ, can be found by differentiating Eq. (3A). Solving for $d\theta_1$ and simplifying, we obtain

$$d\theta_1=\frac{(1-\cos^2\gamma\sin^2\frac{1}{2}\theta_1)d\phi_1-\cos\gamma\sin\theta_1 d\gamma}{\sin\gamma}. \tag{4A}$$

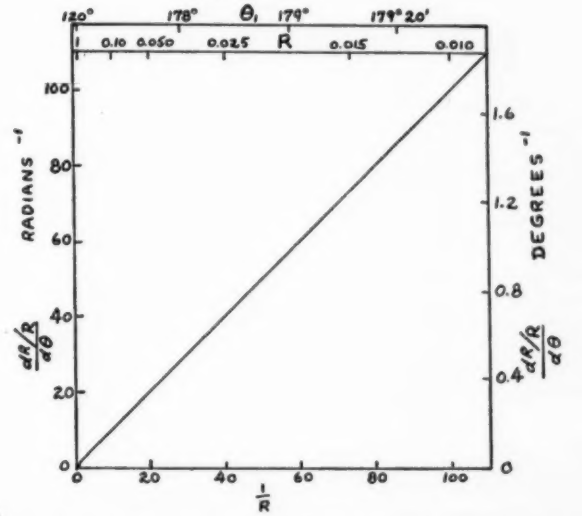


FIG. 11. Plot of the relative error in determining the ratio (R) of two interfacial free energies, per unit error in measuring the dihedral angle (θ_1) between the comparison interfaces.

[16] J. C. Fisher and C. G. Dunn, "Surface and interfacial tensions of single phase solids," National Research Council, Pocono Conference on Imperfections (1950). Discussion by R. Shuttleworth.

[17] R. D. Heidenreich and W. Shockley, "Study of slip in aluminum crystals by electron microscope and electron diffraction methods," Report of Bristol Conference, Phys. Soc. (1948).

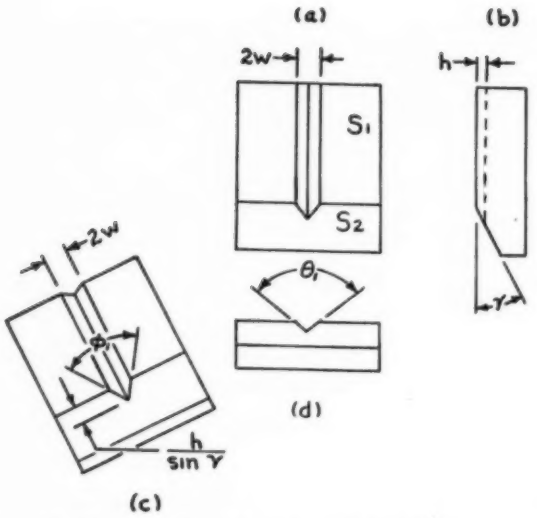


FIG. 12. Diagram illustrating the relationships between the dihedral angle (θ_1) of a groove in a specimen surface and the angle (ϕ_1) between the groove traces on an oblique section, when the groove is perpendicular to the intersection of the specimen surface and the oblique section. (a) Projection parallel to specimen surface. (b) Projection perpendicular to specimen surface and oblique section. (c) Projection parallel to oblique section. (d) Projection perpendicular to groove.

If the probable errors in measuring ϕ_1 and γ are about equal, we may set them both equal to a quantity dm, which represents the error in measuring angles.

$$dm = |d\phi_1| = |d\gamma|.$$

For all values of θ_1 and γ that have physical meaning the coefficients of $d\phi_1$ and $d\gamma$ in Eq. (4A) have positive signs. Hence, on substituting dm for $|d\phi_1|$ and $|d\gamma|$ the sign of the second term must be reversed, in order to reflect the cumulative error in θ_1 that may result when the errors in measuring ϕ_1 and γ are in opposite directions. The error in determining θ_1 per unit error in measuring the angles is then

$$\left|\frac{d\theta_1}{dm}\right| = \frac{1+\cos\gamma\sin\theta_1-\cos^2\gamma\sin^2\tfrac{1}{2}\theta_1}{\sin\gamma}, \qquad (5A)$$

and the relative error in the interfacial free energy ratio is

$$\left|\frac{dR/R}{dm}\right| = \frac{\tan\tfrac{1}{2}\theta_1}{2\sin\gamma}\left(1+\cos\gamma\sin\theta_1-\cos^2\gamma\sin^2\frac{\theta_1}{2}\right). \qquad (6A)$$

In order to find the optimum sectioning angle, γ, for any estimated value of θ_1 to be measured, it is necessary to find the value of γ for which $|d\theta_1/dm|$ is a minimum. Differentiating Eq. (5A) with respect to γ and setting the result equal to zero, we find

$$\frac{d|d\theta_1/dm|}{d\gamma} = -\frac{\sin\theta_1+\cos\gamma(\cos^2\tfrac{1}{2}\theta_1-\sin^2\gamma\sin^2\tfrac{1}{2}\theta_1)}{\sin^2\gamma} = 0. \qquad (7A)$$

Equation (7A) may be solved by successive approximations, to find the value of the sectioning angle γ that gives optimum precision in determining a dihedral angle θ_1. Figure 13 shows the locus of Eq. (7A); the lower branch of the curve indicates the optimum section angles. When a dihedral angle less than about 159° is to be measured, no increase in precision can be obtained by using an oblique cross section, unless the angle between the specimen surface and the cross section (γ) can be measured more precisely than that between the traces of the groove sides. Figure 14 shows the variation of $|dR/R/dm|$ with γ, for several values of θ_1. It can be seen that no worth-while increase in precision is obtained by oblique sectioning unless the dihedral angle to be determined is very near 180°.

Previous experiments indicated that thermal etching of copper in the presence of lead vapor would produce twin boundary grooves with a dihedral angle of about 178.4°. Figures 13 and 14 show that measurements on a cross section at an angle of 10° from the surface provide optimum precision in measuring groove angles of this size. With the dihedral angle actually found (179.04°) and the sectioning angle actually used (8.79°), the precision was near optimum. When this sectioning angle is used, the error in determining the interfacial free energy ratio, per unit uncertainty in the measured angles, is about one-fourth the error that would be obtained using measurements on a section perpendicular to the surface.

The lower curve of Fig. 14 shows the variation of $|(dR/R)/dm|$ with γ, corresponding to the mean dihedral angle predicted at the grooves where grain boundaries intersect the specimen surface. Observation of these groove traces on the oblique section at 8.79° leads to an error about four times as large as would be obtained if the measurements were made on a cross section perpendicular to the surface. However, the contribution of the error in the grain boundary groove angle to the error in σ_I/σ_G is still smaller than the contribution due to uncertainty in the twin boundary groove angle, as shown by the corresponding curves of Fig. 14. Hence, no significant increase in the accuracy of the determination of σ_I/σ_G would be obtained by measuring the grain boundary groove traces on a section perpendicular to the specimen surface. Actually, an error in the sectioning angle γ changes the ratios $\sigma_I/\sigma_{Cu-Pb(v)}$ and $\sigma_G/\sigma_{Cu-Pb(v)}$ in the same direction, by an amount nearly proportional to their values, so that this source of error nearly vanishes from the ratio σ_I/σ_G.

Calculation of Dihedral Angles from Measurements on an Oblique Section

In determining the value of the dihedral angle between the sides of a groove in the specimen surface, Eq. (3A) may not be used, since in general the groove is not perpendicular to the intersection between the specimen surface and the section on which the observations are made. In the following analysis it is assumed that the grain and twin boundaries are perpendicular to the specimen surface. Observations of perpendicular cross sections, of the positions of boundary traces on opposite sides of sheet specimens, and of the orientations of grains and their twins, all indicated that nearly all the boundaries are quite close to being perpendicular to the sheet surface in specimens of the type used in these experi-

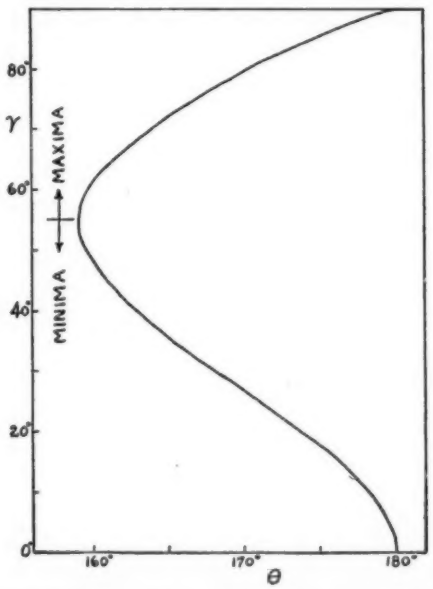


FIG. 13. Plot of values of the section angle γ at which $|d\theta_1/dm|$ is a minimum or a maximum, as a function of the dihedral angle θ_1.

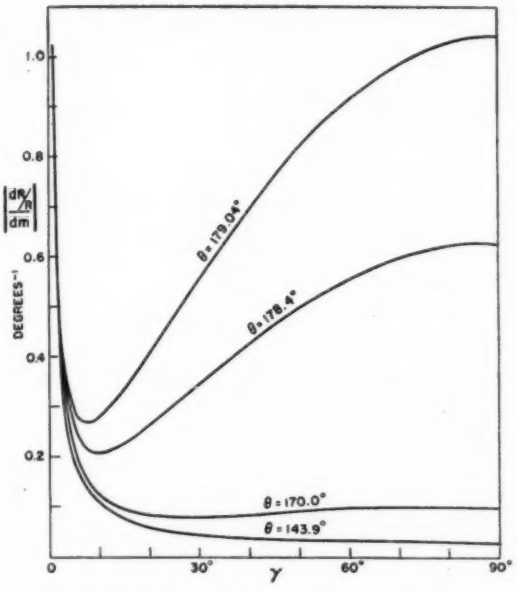


FIG. 14. Plot of the relative error in determining an interfacial free energy ratio, per unit error in measuring ϕ and γ, as a function of the section angle γ.

ments. It is further assumed that the interfacial free energy of the copper/lead (vapor) interface is independent of the orientation of the copper surface, so that the bisectors of the groove angles are also perpendicular to the specimen surface.

Consider the specimen shown in Fig. 15, containing a groove with dihedral angle θ, making an angle α with the intersection of the specimen surface and the oblique section. β is the angle between the boundary trace on the oblique section and the intersection between this section and the specimen surface. ϕ is the angle between the traces of the groove sides on the oblique section. The relationship between the angle ϕ measured on the oblique section and the dihedral angle θ may be expressed by the formula given by Harker and Parker.[18] In the present notation this takes the form

$$\tan\phi=\frac{2\sin\theta\sin\gamma}{\cos^2\gamma(\cos 2\xi-\cos\theta)+2\cos\theta}. \tag{8A}$$

[18] D. Harker and E. R. Parker, Trans. Am. Soc. Metals **34**, 156 (1945).

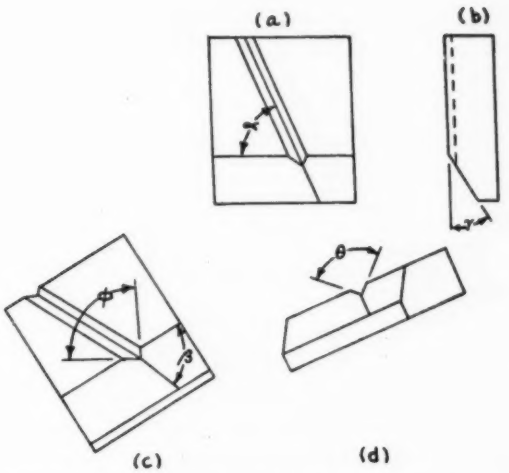


FIG. 15. Diagram illustrating the angles formed by groove and boundary traces on an oblique section. (a) Projection parallel to specimen surface. (b) Projection perpendicular to specimen surface and oblique section. (c) Projection parallel to oblique section. (d) Projection perpendicular to groove.

ξ is the angle between the bisector of the dihedral angle and the projection, on a plane perpendicular to the groove, of the normal to the oblique section.

Solving Eq. (8A) for θ and simplifying the notation, we find

$$\cos\theta=\frac{-a^2cd\pm b[b^2+a^2(d^2-c^2)]^{\frac{1}{2}}}{a^2d^2+b^2}, \tag{9A}$$

where

$$a=\tan\phi,$$

$$b=2\tan\gamma\sec\gamma,$$

$$c=\cos 2\xi=\frac{\cot^2\gamma(1+\cos^2\gamma\tan^2\beta)-1}{\cot^2\gamma(1+\cos^2\gamma\tan^2\beta)+1},$$

$$d=2\sec^2\gamma-1.$$

Equation (9A) permits evaluation of the dihedral angle θ from the values of ϕ and β observed on a section that makes an angle γ with the specimen surface. When there is uncertainty concerning which sign of the radical should be used, substitution, in Eq. (8A), of the two values of θ obtained will indicate which is the correct value.

Crystallography and Interfacial Free Energy of Noncoherent Twin Boundaries in Copper

R. L. FULLMAN*
General Electric Research Laboratory, Schenectady, New York
(Received October 23, 1950)

The crystallography and interfacial free energy of noncoherent twin boundaries in copper have been investigated. It is found that the noncoherent twin boundary is approximately parallel to a {113} plane of one crystal and to a {335} plane of the other. The ratio of the interfacial free energy of noncoherent twin boundaries to the average grain boundary free energy in copper is found to be 0.80±0.015. Two measurements by a second method confirm the magnitude of this value.

A DOMINANT feature of the microstructures of most annealed metals and alloys with a face-centered cubic crystal structure is the presence of many straight-sided crystals that have a twinned orientation relative to the grains in which they appear. These crystals are called annealing twins. The boundary between twinned crystals frequently coincides with the twinning plane, or reflection plane relating the crystal orientations. Since neighboring atoms can match up perfectly across a twin boundary that coincides with the {111} twinning plane of face-centered cubic crystals, it has been suggested that a boundary with this orientation be called a "coherent twin boundary."[1] Twin boundaries that do not coincide with the twinning plane do not permit exact matching of neighboring atom positions across the boundary, and such boundaries were termed "noncoherent twin boundaries." Greninger[2] has reported that the noncoherent twin boundary is a {110} plane that is common to both the crystal and its twin. However, the small angle frequently observed between two segments of a noncoherent twin boundary trace, as in Fig. 1, does not appear to be consistent with planes of this type. The smallest angle between adjacent {110} planes that are common to twin crystals is 60°.

CRYSTALLOGRAPHY OF NONCOHERENT TWIN BOUNDARIES

A technique was developed that permits determining the crystallography of interfaces that appear in ordinary microcrystals, and eliminates the necessity of preparing crystals with boundaries of sufficient length to permit sectioning and examination by ordinary techniques. A metallograph accessory with the design shown in Fig. 2 was constructed. Part *A* is a ring machined to fit accurately into the opening in the top of the metallograph table. A slot in the flange fits snugly over a pin in the metallograph table, so that part *A* can be accurately positioned on this table. Part *A* also carries

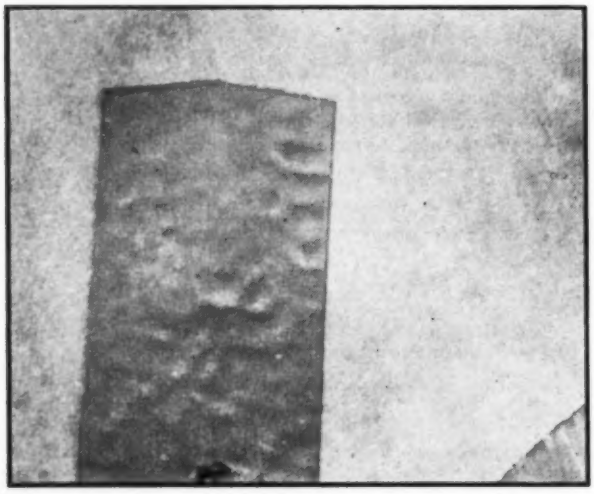


FIG. 1. Compression rolled copper sheet annealed for 18 hours at 1040°C. Traces of noncoherent twin boundaries across the end of a twin that ends within a grain. 160×. Ammonium peroxide etch.

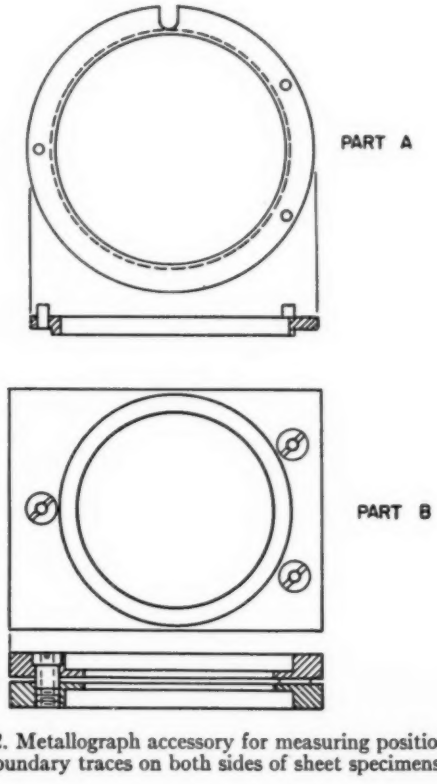


FIG. 2. Metallograph accessory for measuring positions of boundary traces on both sides of sheet specimens.

* This paper is based upon a part of a dissertation submitted to the faculty of the Engineering School of Yale University in partial fulfillment of the requirements for the degree of Doctor of Engineering, June, 1950.

[1] R. L. Fullman, J. Appl. Phys. **22**, 448 (1951) (preceding paper).

[2] A. B. Greninger, Trans. Am. Inst. Mining Met. Engrs. **122**, 74 (1936).

three short steel pins located as shown in the drawings. Part B is a clamp for holding sheet specimens. The two halves of part B are held together by three hollow screws which fit snugly over the pins on part A. By means of these pins, the clamp may be located on the metallograph table either face up or face down, the two positions being related by a 180° rotation about an axis in the plane of the specimen held in the clamp. A strip of sheet metal to be examined was prepared for metallographic observation on both sides. Irregularities along the edges of Scotch tape stretched across the specimen were observed from both sides, and were used as reference marks. Positions of reference marks and points in the metal structure were located by means of the longitudinal and transverse micrometer screws of the metallograph table motion, which form a set of orthogonal axes.

The measurements were analyzed as follows. Let the coordinates of a reference point observed on the edge of the Scotch tape be (x_F', y_F') when viewed from the front of the specimen, and (x_B', y_B') when viewed from the back. Then the equation of the axis, about which the clamp is rotated when it is turned over, may be put in the form $y=mx+b$, where the constants m and b are given by:

$$m=-\frac{x_B'-x_F'}{y_B'-y_F'}.$$
$$b=\frac{(x_B'^2-x_F'^2)+(y_B'^2-y_F'^2)}{2(y_B'-y_F')}. \tag{1}$$

Then for any point (x_B, y_B) found on the back of the specimen, the corresponding point viewed from the front is located at (x_{BF}, y_{BF}) given by:

$$x_{BF}=\frac{2m(y_B-b)+x_B(1-m^2)}{m^2+1}.$$
$$y_{BF}=\frac{(m^2-1)y_B+2mx_B+2b}{m^2+1}. \tag{2}$$

The angle between a trace and a reference direction, such as the X axis, can also be measured on both the front and back of the specimen. If θ_B is the angle between a given trace and the X axis, as measured on the back of the specimen, and θ_{BF} is the angle that would be measured if this trace could be viewed, through the specimen, from the front, then

$$\theta_{BF}=2\theta_{AX}-\theta_B, \tag{3}$$

where θ_{AX} is the angle between the axis of rotation of the clamp and the X axis. Its value is $\tan^{-1}m$.

Specimens of compression rolled *OFHC* copper were annealed at a sufficiently high temperature to ensure that most of the grains extended completely through the sheet. A specimen was mounted in the clamp as described above and examined on the metallograph at a visual magnification of about 150×. Grains were located which revealed coherent twin boundary traces in at least two directions, as well as one or more noncoherent twin boundary traces. The angles between the coherent and noncoherent twin boundary traces and the X axis were measured, and the positions of a sufficient number of points along these traces were measured to permit defining the shape of the twins in the grain. The readings of the micrometer screws were estimated to the nearest 0.0001 in. Measurements were made on both sides of the specimen, and several reference points along the Scotch tape adjacent to the specimen were located from both sides.

TABLE I.

Point	y_F	x_F	y_B	x_B	y_{BF}	x_{BF}
1	0.2113	0.3238	0.4401	0.1688	0.2076	0.3314
2	0.2080	0.3121	0.4241	0.1588	0.2037	0.3129
3	0.2073	0.3121	0.4246	0.1578	0.2026	0.3130
4	0.2101	0.3248	0.4409	0.1685	0.2071	0.3320
5	0.2123	0.3137	0.4228	0.1630	0.2081	0.3131
6	0.2154	0.3165	0.4250	0.1660	0.2102	0.3152
7	0.2278	0.3032	0.4068	0.1746	0.2245	0.3021
8	0.2150	0.3112	No steps on back			
A	0.7817	0.3606	0.2703	0.7190		
B	0.7974	0.3699	0.2735	0.7367		
C	0.8055	0.3635	0.2655	0.7418		
D	0.2463	0.7494	0.8205	0.3489		
E	0.2438	0.7517	0.8237	0.3473		

	θ_F	θ_B	θ_{BF}	$\bar{\theta}$	d	ψ
1–2, 3–4	14.4°	55.7°	14.3°	14.4°	0.0048″	76.2°
2–3	110.1°	320.0°	110.0°	110.1°	0.0008″	87.6°
7–8	122.7°	307.2°	122.8°	122.7°	0.0024″	83.0°
5–6	41.2°	29.5°	40.5°	40.9°	0.0030″	81.3°
X-axis	28.2°	0				
Rotation axis	35.0°	6.8°				

The equation of the axis of rotation was computed by means of Eq. (1). Using the average values of m and b found from the positions of the reference marks, the positions of points in the structure observed from the back (x_B, y_B) were used to calculate the positions (x_{BF}, y_{BF}) where these points would be located if they could be viewed, through the specimen, from the front. These calculations were made by means of Eq. (2). A large scale plot was made of the structure as viewed from the specimen front. From this plot the distance (d) between any boundary trace and the projection of the trace of the same boundary on the opposite side of the specimen could be measured. The dihedral angle (ψ) between the boundary and the specimen surface is given by Eq. (4).

$$\tan\psi=h/d, \tag{4}$$

where h is the specimen thickness. The values of θ and ψ completely define the orientation of a boundary plane. Since at least two directions of coherent twin boundary traces were present in the structure, the orientations of these {111} planes defined the orientations of the grain and the twins. The poles of other planes in the grain and the twins could then be located on a stereographic projection, and compared with the observed positions of the poles of the noncoherent twin boundary planes.

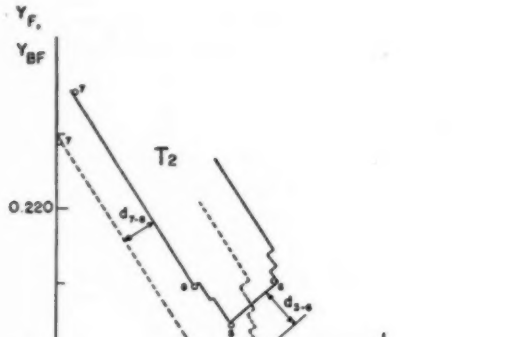


FIG. 3. Locations of twin boundary traces in a grain of annealed copper sheet. Solid lines indicate positions observed on front surface of the sheet. Dashed lines indicate projection of traces calculated from observations on back surface of the sheet.

The technique will be illustrated by a specific example. Table I shows the measured locations of points viewed from the front and from the back of the specimen. The locations of points that define the structure are numbered 1 to 8, and the reference points measured on the Scotch tape are lettered A to E. The specimen thickness, h, was 0.0195 in. Values of θ are tabulated for the boundary traces that extend between various points as indicated. The average values of the constants m and b, in the equation of the axis of rotation of the clamp, were found from the positions of the reference points to be 0.6993 and 0.1490, respectively. Using these values in Eq. (2), values of x_{BF} and y_{BF} were calculated for the various structural points observed on the back of the specimen (x_B and y_B). Values of θ_{BF} were calculated from Eq. (3). Figure 3 is a plot of these data, where the shape of twin T_1 is defined by the points 1, 2, 3, 4 and the shape of twin T_2 is defined by the points 5, 6, 7, and 8. The distance between each boundary trace on the specimen front and the projection of its position on the specimen back was measured, and the corresponding values of ψ were calculated using Eq. (4).

Figure 4 is a stereographic projection of the planes shown in Fig. 3. The poles of the planes of the grain and the twins that fall closest to the observed positions of the poles of the noncoherent twin boundary planes are plotted. Figure 5 is a cumulative plot, rotated into a standard projection position, of all the measurements made. It will be observed that the locations of poles of the noncoherent boundaries are clustered around positions between a {113} pole of one crystal and a {335} of the other.

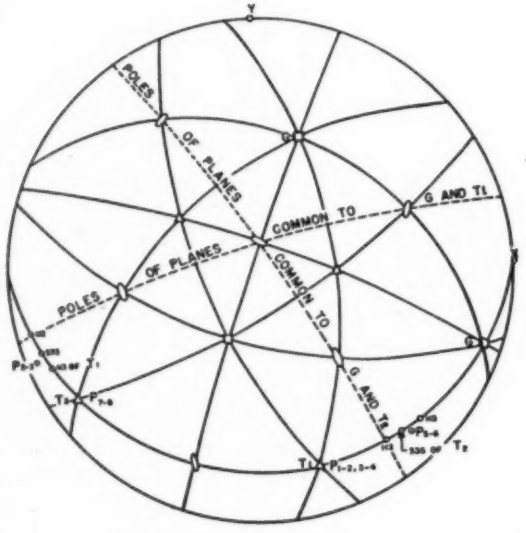


FIG. 4. Stereographic projection of the twin boundaries shown in Fig. 3. $P_{1-2,\ 3-4}$, and P_{7-8} are the poles of the coherent twin boundaries of twins T_1 and T_2. P_{2-3} and P_{5-6} are the poles of noncoherent twin boundaries of twins T_1 and T_2.

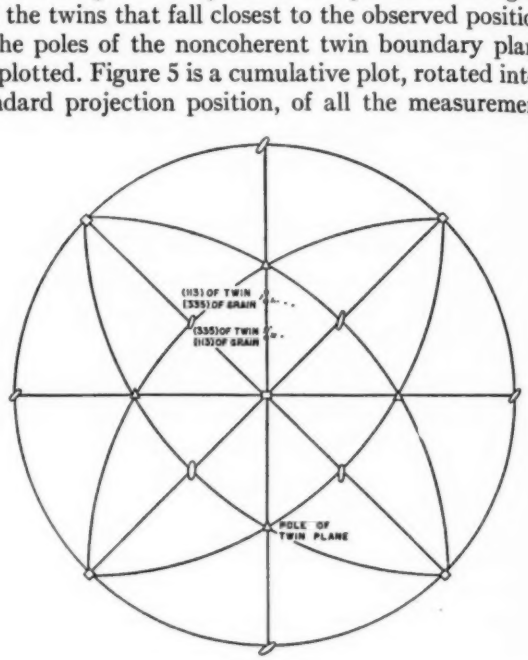


FIG. 5. Stereographic projection of poles of 16 noncoherent twin boundary planes (solid dots) with reference to a standard projection of the grain orientation.

The observation of such a nonrational boundary plane certainly demands a careful scrutiny of the measurements, particularly in view of the fact that a new and untried technique was used. One objection that may be raised to the technique is that it is necessary to assume that the boundaries concerned pass completely through the specimen thickness as a plane, without the presence of any other "steps" in the interface. However, it would be expected that the occurrence of such steps would lead to a random scatter in the observations, rather than a systematic error. That such steps do occasionally occur was indicated by the fact that, for a few of the twin groups observed, the dihedral angle between {111} planes apparently did not coincide with 70° or 110°. These structures were not used.

RATIO OF THE INTERFACIAL FREE ENERGY OF NONCOHERENT TWIN BOUNDARIES TO THAT OF GRAIN BOUNDARIES IN COPPER

The angles formed at the intersections of noncoherent twin boundaries and grain boundaries were measured for structures similar to that shown in Fig. 6. In analyzing the data it was assumed that surface tensions were equilibrated only in a direction parallel to the plane of the noncoherent twin boundary, since the predominantly straight traces of the noncoherent twin boundary indicate that it lies in a "free energy cusp" with respect to boundary orientation. It was assumed that the two grain boundaries have the same free energy. This assumption does not lead to serious error if the ratio of the noncoherent twin boundary free energy to that of grain boundaries is relatively large, as was suspected. Under these assumptions the ratio of the free energy of the noncoherent twin boundary to that of the grain boundaries is given by the equation:

$$\sigma_{t'}/\sigma_G = -(\cos\theta_1 + \cos\theta_2), \qquad (5)$$

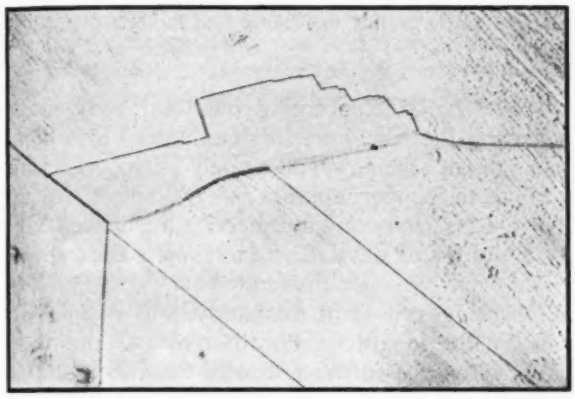


FIG. 6. Copper sheet annealed 18 hr at 1040°C. Intersection of a noncoherent twin boundary with a grain boundary. 400×. Potassium dichromate etch.

where $\sigma_{t'}$ is the interfacial free energy of a noncoherent twin boundary and σ_G is the interfacial free energy of a grain boundary. θ_1 and θ_2 are the dihedral angles indicated in Fig. 7.

Specimens of *OFHC* copper bus-bar 0.200 in. in thickness were annealed at 500°C for one hour and compression rolled in accordance with one of two rolling schedules, as follows: (1) Compression roll to 0.020-in. thickness, with two intermediate anneals at 500°C. (2) Compression roll to 0.025-in. thickness, with no intermediate anneals. Specimens prepared by either of these schedules were annealed for 18 hr at 1040°C. The angles were measured at 74 intersections of the type shown in Fig. 7. No significant difference was found in the results obtained on the two types of specimen.

The ratio $\sigma_{t'}/\sigma_G$ was calculated from Eq. (5), using the observed angles between the boundary traces in place of the dihedral angles. Figure 8 shows the distribution of the ratios found. The mean value found for

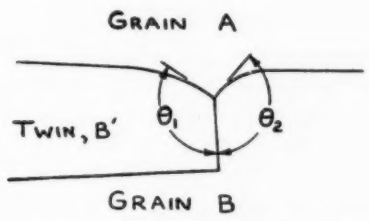


FIG. 7. Diagram showing the angles measured at intersections of noncoherent twin boundaries with grain boundaries.

the ratio of the interfacial free energy of a noncoherent twin boundary to the average grain boundary free energy is $\sigma_{t'}/\sigma_G = 0.80 \pm 0.015$.

Data collected from several sources[3] indicate that the average grain boundary free energy in copper is about 600 ergs cm^{-2}. Accepting this value, the interfacial free energy of noncoherent twin boundaries in copper is about 480 ergs cm^{-2}.

During observation of the dihedral angles of grooves formed at coherent twin boundaries and grain boundaries on thermal etching copper sheet for 100 hours at 945°C in lead vapor,[1] two intersections of noncoherent twin boundaries with the copper/lead vapor interface were found. For these two intersections the dihedral angles ($\theta_{t'}$) observed at the grooves were 155.9° and 152.5°. The mean value found for the dihedral angle (θ_G) at grain boundary grooves was 142.0°±1.4°. For the two values of $\theta_{t'}$ observed, we find:

$$\frac{\sigma_{t'}}{\sigma_G} = \frac{\cos(\theta_{t'}/2)}{\cos(\theta_G/2)} = 0.73,\ 0.64.$$

The values confirm the order of magnitude of the value found by measurements of the angles formed at intersections of noncoherent twin boundaries with grain boundaries.

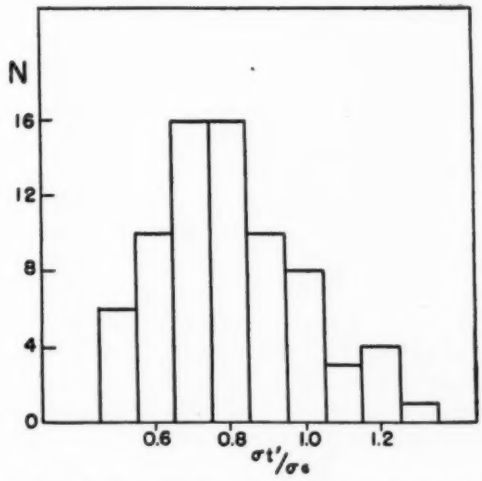


FIG. 8. Distribution of values found for the ratio ($\sigma_{t'}/\sigma_G$) of the interfacial free energy of a noncoherent twin boundary to that of grain boundaries in copper.

[3] J. C. Fisher and C. G. Dunn, National Research Council, Pocono Conference on Imperfections (1950). Discussion by R. Shuttleworth.

CONCLUSIONS

The crystallography and interfacial free energy of noncoherent twin boundaries in copper have been investigated. It is found that the noncoherent twin boundary is approximately parallel to a {113} plane of one crystal and to a {335} plane of the other. The ratio of the interfacial free energy of noncoherent twin boundaries to the average grain boundary free energy in copper is found to be 0.80 ± 0.015. Two measurements by a second method confirm the magnitude of this value.

ACKNOWLEDGMENT

The author is indebted to many members of the Metallurgy Division of the General Electric Research Laboratory for assistance, and in particular to J. C. Fisher and G. W. Sears. The investigation was supported by the AEC.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 4 APRIL, 1951

Leak Location by Radioactive Gases in Buried Pipes*

ANDREW GEMANT, EDWARD HINES, AND E. L. ALEXANDERSON
Research Department, The Detroit Edison Company, Detroit, Michigan
(Received September 29, 1950)

This paper describes basic considerations and experiments on leak location by means of radioactive tracer gases in underground pipes. C-14 labelled carbon monoxide and radon were used as tracers. The principles of the method are described. Suitable techniques for preparing and detecting the tracer gas have been worked out. Calculations on the spreading of a column of tracer gas in a pipe follow. Leak location on two small scale installations was carried out experimentally.

INTRODUCTION

THE principle of locating, by means of radioactive gases, leaks in closed containers can be applied to any closed system and has thus a wide range of applicability. The chief field of application appears to be containers that are not easily accessible, such as buried pipes. The method deserves serious consideration for pipe-type high voltage cable lines.

There are several existing methods for locating leaks in pipes, such as that developed by J. D. Piper,[1] but none of the existing methods is applicable to all conditions. The radioactive method has the advantages of simplicity, operating on a purely qualitative principle, great sensitivity, and the need of little equipment.

One of the chief problems is the proper choice of the tracer element and the gas containing the tracer atom. The chief requirement the tracer must fulfill is its detectability at a concentration sufficiently low so that it does not constitute either a health hazard or a danger to the equipment to be tested. The ideal tracer should have a moderately long (few days to few weeks) half-life.

A fairly suitable tracer (except for its long half-life) was found in the form of C-14 which can be converted into both carbon dioxide and carbon monoxide. Tests showed that a microcurie of such a gas can be detected in concentrations of 10^{-4} μc/cc. According to present knowledge[2] the tolerance concentration of C-14 in air in case of permanent exposure (worst condition) is 5×10^{-5} μc/cc. Concerning the total body intake, the amount that could possibly be retained by a person in our present scheme would be only microcuries which according to Skipper[3] appears safe. Although the half-life of C-14 is 5360 years, its effective life in the body is only a matter of days. The gas chosen thus appears satisfactory from a health standpoint.

A second gas, used in this study, was radon. After its decay the long-lived Pb-210 remains, and if the original activity used was not more than about 100 μc, the residual activity will be less than 0.05 μc. Actually less than this will stay in the pipe, if the pipe is flushed after each test. Radon might be replaced by another gas if such should become available. Xenon-127, for instance, with half-life of 34 days and 0.9-Mev gammas, might be suitable for the purpose.

The present study must be considered as a first step along the lines mentioned. Experimentation was carried out in the Laboratory and in two short pipes in the field, but not in long buried pipes so far. The material to be presented in the following pages is somewhat heterogeneous, involving physical, chemical, and engineering aspects. The results of this study are presented under the following headings: principles of leak location; preparation of the tracer gas; detection of the tracer gas; introduction and subsequent spreading of the gas in the pipe; location of leak.

I. PRINCIPLES OF LEAK LOCATION BY RADIOACTIVE TRACERS

Among the various possible principles, two were studied in this investigation.

* A preliminary report of this study was presented at the 1950 Fall Conference on Electrical Insulation of the National Research Council at Pocono Manor, Pennsylvania.

[1] J. D. Piper, Trans. Am. Inst. Elec. Engrs. **67**, 10 (1948).

[2] K. Z. Morgan, J. Phys. Colloid Chem. **51**, 984 (1947).

[3] Skipper, White, and Bryan, Science **110**, 306 (1949).

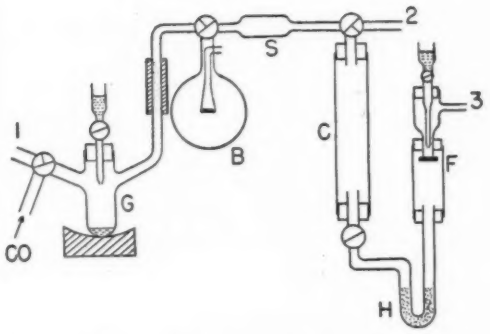


FIG. 1. Diagram of apparatus for generating and detecting labelled carbon monoxide.

1. A radioactive gas having *soft radiation* (CO) is introduced at a suitable point into a buried pipe. If the pipe has a leak, the tracer will move in that direction, and, reaching the leak, will escape into the ground where it will diffuse out in all directions, but preferentially along the pipe where the soil may be looser. At certain spacings bore-holes in the ground over the line are prepared and the air from the soil is pumped into the detecting apparatus. The bore-hole with positive indication then identifies the location of the leak.

2. A radioactive gas emitting *hard radiation* (Rn) is introduced into the pipe. As it spreads out in the direction of the leak, its progress may be detected by surveying the line with a counter. At the leak the gas escapes and from that point on no further radiation along the line is detected. An alternative of this principle consists in measuring the activity merely in the immediate vicinity at both sides from the point of introduction. The higher activity indicates the direction of flow and of the leak from the point in question. In this manner a pipe, accessible by manholes, may be investigated and the leaky man-hole section sought. Having ascertained the latter, Principle 1 may be used for narrower location of the leak.

II. PREPARATION OF THE TRACER GAS

1. Carbon monoxide

Flask *G* (Fig. 1) (volume 20 cc) served for generating CO_2 from radioactive barium carbonate and concentrated sulfuric acid. Usually a charge of about 0.01 g $BaCO_3$, containing from 0.1 to 0.2 mc of C-14 was employed. A small heater served to expel the absorbed gas from the liquid phase. The gas was flushed by non-radioactive CO into the tungsten-filament light bulb *B* (volume 1000 cc), the method suggested by Kummer[4] being used for conversion of CO_2 into CO. In order to remove traces of O_2, the gas, prior to its entering *B*, was passed through a tube containing fine copper wire, heated to about 500°C. The bulb was operated at 60 volts and 3 amperes for 20 hours. About 95 percent conversion efficiency is obtainable.

[4] J. T. Kummer, J. Am. Chem. Soc. **69**, 2239 (1947).

FIG. 2. Diagram of quartz-platinum cell for liberating radon from radium.

2. Radon

The generator (Fig. 2) is a quartz container with a platinum crucible inside; the top of the quartz cell is closed after the crucible is placed inside. The crucible contains a small amount (0.1 mc) of radium salt with four times its weight of a mixture of sodium and potassium carbonate. Placed upon a Meker burner, the mix melts and the occluded Rn is liberated. One end of the cell is connected to the tank in which the gas is to be collected and which is previously evacuated. By opening the stopcock, the gas is swept into the container. The other stopcock serves for flushing the cell and transferring the maximum possible Rn into the container.

III. DETECTION OF THE TRACER GAS

Simple immersion of a thin-window counter into a gas space containing the tracer produces relatively low counts. Accordingly, a technique was devised involving enrichment of the CO in a small and light filter. Such a filter, brought close to the mica window of a Geiger-Mueller counter, radiates into the counter under very favorable geometric conditions. Since it is not easy to absorb carbon monoxide by a filter, the latter was first converted into carbon dioxide which is easily absorbed by a suitable filter.

The reason that CO and not CO_2 was used as the tracer gas is that the latter is readily absorbed by soil, as special experiments have shown; thus, it would not spread out sufficiently in the ground after it had escaped from a leak.

The conversion of CO into CO_2 prior to filtering is accomplished by passing the gas through a tube H (Fig. 1) containing Hopcalite,* a mixture of chiefly copper and manganese oxides; this material converts CO into CO_2. The filter *F* (fiberglass mat, diameter 1.3 cm, thickness 0.05 cm) is cemented to the end of a glass tube. The filter contains alcoholic potassium hydroxide solution. The filter is kept wet while the gas is passing through it; this is accomplished by means of the alcohol dropper shown. The gas to be analyzed is drawn through the filter for about 10 minutes, a pump being connected to opening 3. Toward the end of this process the filter is permitted to dry, formation of K_2CO_3 being noticeable by its white color.

In order to determine the sensitivity of the filter detector, a known amount of tracer was drawn from

* Mine Safety Appliances Company, Pittsburgh, Pennsylvania.

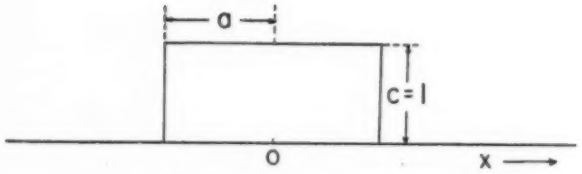


FIG. 3. Diagram of a tracer gas column in a pipe.

the bulb B (Fig. 1) into the sampling tube S, and diluted with air in the cylinder C so that the mixture corresponded to a definite concentration of tracer in air. Two μc of CO in a concentration of 10^{-3} μc/cc produced after filtering about 10,000 counts per minute. Removal by means of a suitable absorbent all extraneous CO_2 from the tracer-air mixture prior to its passage through Hopcalite is expedient.

The efficiency of CO collection by the alkaline filter was obtained from counts taken at a known distance of the filter from the counter window. The expected counts for the case that the entire 2 μc was fixed by the filter, can be calculated[5] from the weight of the filter (18 mg); the actual counts indicated an efficiency of about 24 percent. This low efficiency was expected in view of the thinness of the filter, permitting a large fraction of the CO_2 to pass unabsorbed.

The effect of the soil in using this technique was shown by permitting a sample to stand in contact with soil in cylinder C for 16 hours: a count of 800 per min was obtained in a particular case; a control test without soil gave the same counts. Hence, CO can be used in underground installations as was to be expected from the low solubility data for CO in water (Ostwald absorption coefficient 0.02 as compared with 0.9 for CO_2).

IV. INTRODUCTION AND SPREADING OF TRACER GAS IN PIPE

The question arises how much tracer gas is required if a leak in a given length of pipe is to be located. The travel of a column of tracer gas down the pipe is due to the drop of pressure at the leak, while the original pressure is maintained by pressure tanks connected to the pipe. In addition to this linear displacement of the column, the latter will spread because of diffusion. In the radioactive technique this spreading of the column by diffusion is desirable, since in this manner the tracer gas reaches the leak faster and the time required until the tracer can be detected is reduced.

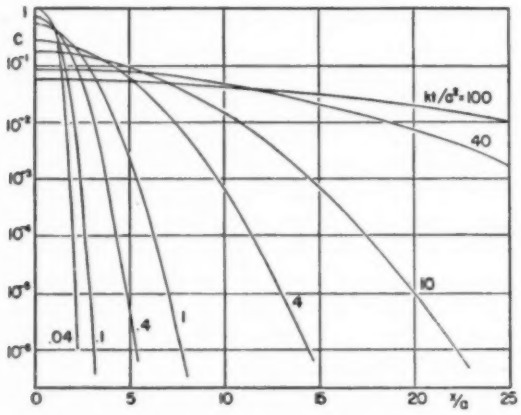


FIG. 4. Spread of tracer gas column for various values of kt/a^2.

[5] Edward Hines and Andrew Gemant, Science **110**, 19 (1949).

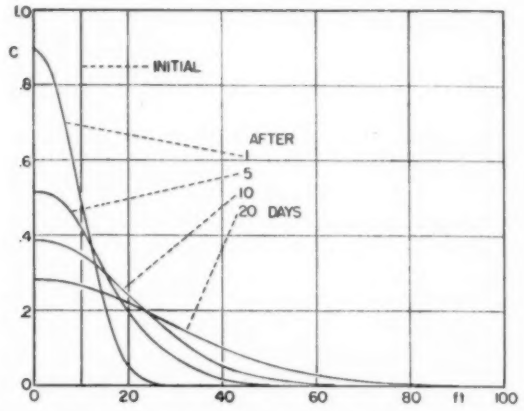


FIG. 5. Spread of originally sharply bounded column ($a=10$ feet) after 1, 5, 10, and 20 days for carbon monoxide ($k=0.2$ $-cm^2/sec$).

Consider a gas column of length $2a$ in a pipe, any point along which is designated by x; the center of the column coincides with $x=0$ (Fig. 3). This column contains a tracer gas at a concentration designated by unity. The boundary condition is: for the time $t=0$, the concentration $c=1$ for $-a<x<a$, and $c=0$ outside this range. The concentration for any time and any location can be calculated by starting out from a general equation, as given, for instance, by Carslaw and Jaegar.[6] The result of this calculation is:

$$c=0.5\left[\operatorname{erf}\frac{(x/a)+1}{2(kt)^{\frac{1}{2}}/a}-\operatorname{erf}\frac{(x/a)-1}{2(kt)^{\frac{1}{2}}/a}\right]. \qquad (1)$$

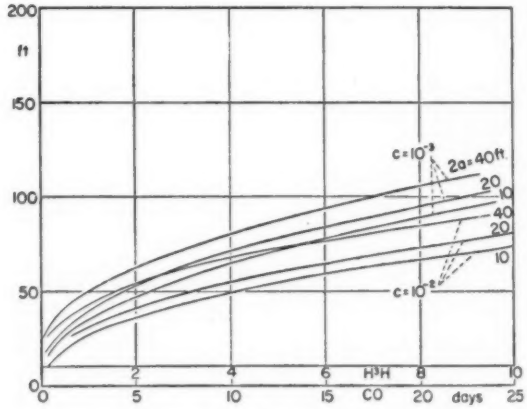


FIG. 6. Rate of spread of tracer column for $2a=10$, 20, and 40 feet, for carbon monoxide and tritiated hydrogen gas.

[6] H. S. Carslaw and J. C. Jaeger, *Conduction of Heat in Solids* (Clarendon Press, Oxford, England, 1947), page 34.

In this equation k is the diffusion constant of the tracer gas.

A graphical presentation of Eq. (1) is given in Fig. 4, showing c as a function of x/a; the parameter of the set of curves being kt/a^2. By using this graph, any other desired graph may be drawn. Such a graph is shown in Fig. 5 for carbon monoxide as a tracer, k of which is about 0.2 cm²/sec. It shows the spread of an originally sharply bounded column of 20 ft length ($2a$) after 1, 5, 10, and 20 days.

Another graph is given in Fig. 6. Suppose the original concentration of the radioactive gas is such that 1 or 0.1 percent of it, leaving the leak, could be detected by the method described. It is then of interest to know at what distance from the origin these small concentrations prevail after a number of days. That distance must be added to the displacement of the column due to convection as mentioned earlier. This information is given in Fig. 6 for an original length of 10, 20, and 40 ft; for both carbon monoxide and tritiated hydrogen (H^3H). This latter gas whose diffusion coefficient is

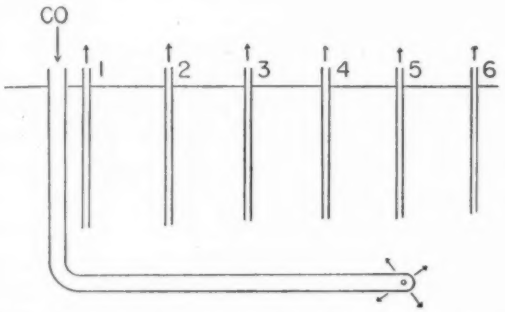


FIG. 7. Diagram of underground pipe installation for leak location by labelled carbon monoxide.

about 0.5 might be used with advantage; its diffusion, also in the soil, is 2½ times faster than that of CO.

From the results obtained it is possible to give a design of a leak test on a longer pipe line. If an installation covers several miles, it might be desirable as mentioned first to ascertain which manhole section is the leaky one. When the leaky section of, say, ½-mile length has been located in that manner, the carbon monoxide technique can be used for a closer location.

The laboratory test showed that 2 μc of CO in a concentration of 10^{-3} μc per cc produced counts of up to 10,000 per min. From this it can be estimated that in a pipe of, say, 2000 ft, perhaps five mc of CO should be used. If the original column is spread over 20 ft, the original concentration (assume, for example, an 8-in. pipe containing three 138-kv cables and having an open cross-sectional area of 0.23 sq. ft) is 40×10^{-3} μc/cc. The rate of travel of this column is about 60 ft per day if the leak is such that the loss of gas at 200 lb/sq. in. pipe pressure is one standard 200 cu. ft tank per day. If the leak is, say, midway, i.e., 1000 ft away from the point of introduction, 15 days are required

TABLE I. Field test with CO, in dry weather.

Time schedule	Gas introduced into pipe	Net counts per min with G-M counter; bore-holes: 1	2	3	4	5	6
First day, A.M.	First tank	—	6	0	31	—	—
First day, P.M.	Second tank	—	0	19	60	1880	24
Second day, A.M.		—	11	11	7	90	—
Second day, P.M.	Third tank	—	36	—	—	—	20
Third day, A.M.		—	0	12	11	265	9

for the column to reach the leak. Figure 5 shows that in 15 days the column has flattened out to six times its original length, and its average concentration is one-sixth of the original. The column has then a tracer concentration of 7×10^{-3} μc/cc, and it takes about 2 days while all the gas escapes from the leak into the ground. The concentration of the escaping gas is about 0.5×10^{-3} μc/cc.

By applying an artificial leak of, say, two tanks per day at the other manhole (not the one at which the tracer was introduced), the rate of progress may be trebled. Instead of 15 days, it would need only 5 days to find the leak. In this fashion only ⅓ of the tracer would escape, but the sensitivity of the method might still be sufficient for finding the leak.

V. LEAK LOCATION IN PIPES

1. Carbon monoxide tracer

Field tests showing the possibilities of leak detection on a pipe buried underground were carried out in the following manner. A pipe, 1 in. in diameter was installed (Fig. 7) in the ground, the horizontal portion being 20 ft long and 5 ft under the ground surface. This portion was jacked from a trench into position in order not to disturb the soil around the pipe; the end of the horizontal part contained a hole, $\frac{1}{16}$ in. in diameter, serving as artificial leak. The soil immediately above the pipe was pure clay, thus presenting unfavorable conditions for the spreading of the gas. In a sandy soil detection should be easier.

Along the 20-ft pipe, ½-in. tubes, 5 ft apart, were driven into the ground, their open lower ends being 1 ft above the pipe. These bore-holes are shown in the figure, and numbered as 1 to 6, No. 5 being over the leak.

For each test about 0.2 mc of CO was produced. The gas was transferred into from two to four 17 cu. ft

TABLE II. Field test with CO, in wet weather.

Time schedule	Gas introduced into pipe	Net counts per min with G-M counter; bore holes: 1	2	3	4	5	6
First day, P.M.	First tank	—	—	—	—	—	—
Second day, A.M.	Second tank	7	8	7	51	—	3
Second day, P.M.	Third tank	6	4	7	58	670	—
Third day, A.M.	Fourth tank	5	5	9	—	—	15
Third day, P.M.		7	—	—	22	51	13

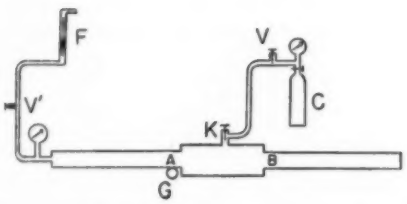


FIG. 8. Diagram of pipe installation to determine direction of leak from a sleeve, using radon.

cylinders, the total pressure in each being brought to 1000 lb/sq. in. These tanks were emptied into the pipe which contained the leak; by means of a reducing valve the pressure in the pipe was kept at 200 lb and the rate of flow was such that each tank emptied in about 3 hours.

From time to time the gas-collecting equipment, consisting of a Hopcalite tube and a filter, was connected to the various bore-holes, and air was pumped out of the soil. The pumping process lasted about 15 min each. Since the Hopcalite retains part of the CO_2, the material must be renewed after each measurement with positive indication.

The results of one particular test are summarized in Table I, giving indications of the time, the radioactive charge emptied into the pipe, and the net counts per min obtained by a recording counter. This test was carried out after a few weeks of comparatively dry weather; the gas permeability of the soil was relatively high.

The results of a second test are shown in Table II. This test was carried out in rainy weather; the moisture content of the soil was higher and the gas permeability of the soil lower than in the first test. The counts obtained were, therefore, generally lower.

It may be seen that the relatively small amount of 0.2 mc dissipating from a leak into the soil is easily detectable by means of the filter technique. There is a maximum of counts at the leak; the counts diminish with increasing distance from the leak along the pipe. Small counts can be found as far as 20 ft away; longer distances along the pipe could not be tested with this installation. A further bore-hole 5 ft away beyond hole 6 in undisturbed ground did not give positive results. The gas obviously travels along the pipe easier than into undisturbed ground. It follows that bore-holes must be spaced not more than 40 ft apart.

TABLE III. Results of first Rn test for directional detection.

Time in hours after introduction of gas	Counts per minute	
	Toward leak	Away from leak
6	180	110
24	110	60

TABLE IV. Results of second Rn test for directional detection.

Time in hours after introduction of gas	Counts per minute	
	Toward leak	Away from leak
2	140	90
3	140	60
5	80	40
6	70	40

2. Radon as tracer

As mentioned in Sec. I, radon was used in order to find the direction of a leak from an accessible point of a pipe.

The test was carried out in an installation shown in Fig. 8. Two 30-ft lengths of 8-in. pipe are joined by a 14-in. sleeve, 9 ft long. The sleeve has a valve K which served for introduction of the radon from a container C. The pressure in the container was brought to 800 lb/sq in., and a needle valve V permitted the slow passage (in about 30 min) of the gas into the pipe the pressure in which was 180 lb.

One end of the pipe was connected to a needle valve V' and a flowmeter F. The needle valve served as an artificial leak permitting a drop of pressure in the pipe of 50 lb/sq. in. in the course of 6 hours.

The counter tube G, of a portable instrument was placed at points A and B at certain times. The readings (background counts deducted) are shown in Table III for one of the tests. In this test about 100 μc of Rn was passed into cylinder C of 3.5 lit.

In a second test, summarized in Table IV, 50 μc of Rn were used in a small cylinder of 400 cc.

In introducing the gas into the sleeve, its spreading toward both sides could not be prevented. After a few hours, however, the counts toward the leak are higher than in the opposite direction. This shows that determination of the leaky section appears feasible by the method indicated.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 4 APRIL, 1951

Diffraction Pattern of Microwaves Near Rods

C. L. ANDREWS
General Electric Research Laboratory, The Knolls, Schenectady, New York
(Received September 29, 1950)

The diffraction patterns of electromagnetic waves in the neighborhood of rods a half-wave long have been calculated from rigorous electromagnetic theory and checked with 8-cm microwaves. This was the one case in which the patterns could be calculated at short distances from the diffracting objects without undue difficulty. The agreement of measurements and exact theory was sufficient to suggest that microwave measurements within a few wavelengths of diffracting objects will be a fruitful approach to diffraction problems. Heretofore most calculations of diffraction patterns at large distances have been based on blind assumptions as to the distribution of intensity near the diffracting apertures and objects.

INTRODUCTION

THE solution of any problem of diffraction of electromagnetic waves by obstacles and apertures is the solution of Maxwell's equations and the equations for the boundary conditions. However, to the author's knowledge, no calculations have been made on the basis of electromagnetic theory of the diffraction patterns of objects and apertures at distances of the order of a wavelength.

Microwaves, because of their convenient length relative to the objects that man can build, have revealed surprising details in the diffraction patterns near the diffracting objects which had not been predicted even approximately. For instance, when a plane polarized electromagnetic wave is incident normally upon a circular aperture sharp peaks of intensity appear along the magnetic diameter at intervals of a wavelength.[1]

Experimental diffraction measurements in the neighborhood of objects of simple geometrical shape have two values: (1) Application to microwave devices. (2) To suggest clues and simplifications in the more general theoretical solution of problems in the diffraction of all electromagnetic waves.

The purpose of the work reported in this paper was to check experimental measurements against electromagnetic theory for the one case in which the pattern had been rigorously derived at finite distances and could be numerically calculated without undue difficulty.

EXPERIMENTAL ARRANGEMENT

The apparatus was that described in a previous paper. To reduce reflections it was set up out of doors on a roof. An 8-cm wave was produced with a triode oscillator.[2] The reflector of the transmitter was 4 ft in diameter. The distance from the source to the diffracting objects was 75 ft. The centers of the parabolic reflector and diffracting rods were 6 ft above the roof. The block of the optical bench which carried the probe was driven by a motor and screw at the rate of two wavelengths per minute and the intensities recorded on a recording galvanometer (General Electric photoelectric recorder) with a sensitivity of 10 μamp full scale and a period of 0.83 sec.

The probing antenna was the crystal itself, a 1N23B sealed-in silicon crystal with the brass tip of the capsule removed so that the probe was 0.2 wavelength long.

A fine twisted lead of No. 24 wire was attached with soft solder to the ends of the capsule of the crystal and extended in the direction of propagation for three meters and guyed by strings and rubber bands. Thence it extended out of the beam perpendicularly to the electric field to the recorder.

The electric field and the axes of the diffracting rods were vertical. The rods were supported by two horizontal strings through two diametrical holes at the ends so that the distance between rods could be quickly adjusted.

The two major sources of error were: (1) The lack of constancy of intensity in the cross section of the incident unperturbed beam. (2) The finite size of the antenna of the probe.

The means of estimating the error due to finite size of probe has been discussed previously.[1] The greatest source of error in the patterns shown in the following graphs was the variation of intensity of the cross section of the unperturbed beam except at points within an eighth-wavelength of a rod where the greatest error was caused by the finite size of the probe.

SINGLE ROD

The radiation pattern of a rod bearing a sinusoidal current has been rigorously derived for finite distances.[3] However, it is observed experimentally that the only rod in which a plane polarized wave, that is polarized parallel to the rod, will excite a simple sinusoidal current is a rod a half-wave long or less. As the rod becomes longer the current distribution approaches a constant over the rod.

From Stratton's expression the field at any point P on a plane perpendicular to a half-wavelength rod through the center of the rod, is parallel to the rod, varies sinusoidally with r/λ, and in amplitude inversely

[1] C. L. Andrews, J. Appl. Phys. **21**, 761 (1950).
[2] C. L. Andrews, Gen. Elec. Rev. **50**, 40 (1947).
[3] J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill Book Company, Inc., New York, 1941), pp. 454–457.

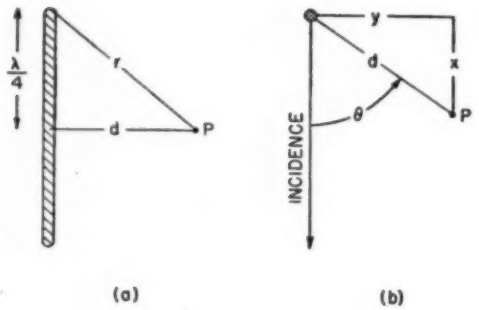


FIG. 1. (a) Cross section including axis of rod and point P. (b) Cross section through point P perpendicular to the rod.

as r, the distance of P from either end of the rod indicated in Fig. 1 (a).

The total field e at point P is the resultant of e_1 and e_2 the fields of the incident plane wave and the reradiated wave from the rod, respectively.

$$e_1=e_0\exp\left[j2\pi\left(\frac{x}{\lambda}-\frac{t}{T}\right)\right] \tag{1}$$

$$e_2=\frac{\lambda}{r}k\exp\left[j2\pi\left(\frac{r}{\lambda}+\alpha-\frac{t}{T}\right)\right]. \tag{2}$$

e_0 is the amplitude of the unperturbed incident wave, x is the component of the distance of P from the center of the rod in the direction of propagation of the incident beam, t is the time, and T the period of the wave, k is an amplitude constant, and α is a phase constant to be determined by the boundary condition $e=0$ at the surface of the rod near the center.

If E_0 and E are the complex rms values of the incident field and the resultant field, then the resultant field in the diffraction pattern is

$$E=E_0\exp\left(j2\pi\frac{x}{\lambda}\right)+\frac{\lambda}{r}k\exp\left[j2\pi\left(\frac{r}{\lambda}+\alpha\right)\right]. \tag{3}$$

At the middle surface of the rod $E=0$, $r/\lambda=\frac{1}{4}$, so that $\alpha=\frac{1}{4}$ and $k=E_0/4$. Thus

$$\frac{E}{E_0}=\exp\left(j2\pi\frac{x}{\lambda}\right)+\frac{\lambda}{4r}\exp\left[j2\pi\left(\frac{r}{\lambda}+\frac{1}{4}\right)\right]. \tag{4}$$

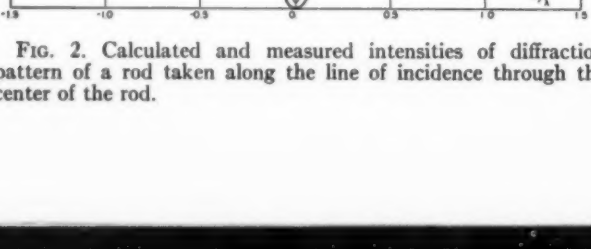


FIG. 2. Calculated and measured intensities of diffraction pattern of a rod taken along the line of incidence through the center of the rod.

If I_0 and I are the intensities of radiation of the incident beam and the resultant diffraction pattern then

$$\frac{I}{I_0}=1+\left(\frac{\lambda}{4r}\right)^2+2\left(\frac{\lambda}{4r}\right)\sin 2\pi\left[\frac{x}{\lambda}-\frac{r}{\lambda}\right]. \tag{5}$$

If θ is the angle measured from the direction of propagation of the incident wave to the line from the center of the rod to P indicated in Fig. 1 (b), then Eq. (5) becomes

$$\frac{I}{I_0}=1+\left(\frac{\lambda}{4r}\right)^2+2\left(\frac{\lambda}{4r}\right)\sin 2\pi\times\left[\left(\left[\frac{r}{\lambda}\right]^2-\left[\frac{1}{4}\right]^2\right)^{\frac{1}{2}}\cos\theta-\frac{r}{\lambda}\right]. \tag{6}$$

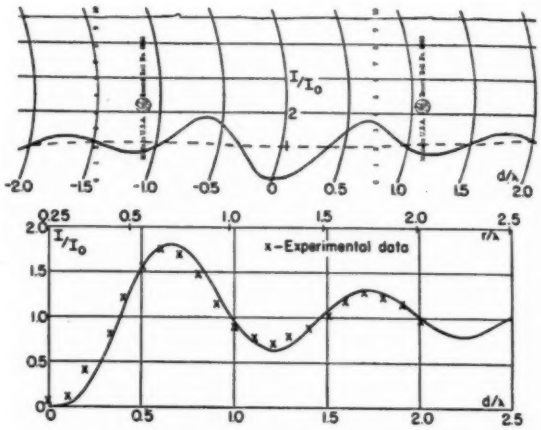


FIG. 3. Experimental record of intensities of diffraction pattern taken along a line perpendicular to the direction of incidence and to the axis of the rod at its center. Comparison of calculations with data taken from the record.

Equation (6) is checked with experiment for three cases: (1) when $\theta=180°$ the two terms in the bracket are of the same sign and the third term of the expression becomes a standing wave term, (2) when $\theta=0°$ and the bracket diminishes rapidly with r/λ, and (3) when $\theta=90°$ and the first term of the bracket is zero.

Figure 2 is for the cases $\theta=0°$ and 180°. The circles indicate experimental data. Figure 3 is of the experimental record for $\theta=90°$ followed by the calculated curve and points of experimental data from the record. The scale of r/λ is indicated along the top. Figure 4 is a plot of amplitude and phase in complex space for $0=90°$. The values of r/λ are indicated by transverse lines along the curve. Figure 5 is a three-dimensional plot of the experimentally recorded intensities in a plane perpendicular to the rod through its center. The maxima and minima lie along a family of parabolas with their common focal point at the rod.

TWO RODS

By the same method the diffraction pattern may be calculated along path AB of Fig. 6 between the centers of two parallel rods a half-wave long whose plane is perpendicular to the direction of propagation of the incident wave. The resultant electric field at any position and time is e made up of three components e_1, e_2, and e_3, the fields of the direct wave and the waves from the two rods. The three fields are expressed by

$$e_1 = e_0 \exp\left(-j2\pi \frac{t}{T}\right) \tag{7}$$

$$e_2 = k\frac{\lambda}{r_2} \exp\left[j2\pi\left(\frac{r_2}{\lambda}+\alpha-\frac{t}{T}\right)\right] \tag{8}$$

$$e_3 = k\frac{\lambda}{r_3} \exp\left[j2\pi\left(\frac{r_2}{\lambda}+\alpha-\frac{t}{T}\right)\right], \tag{9}$$

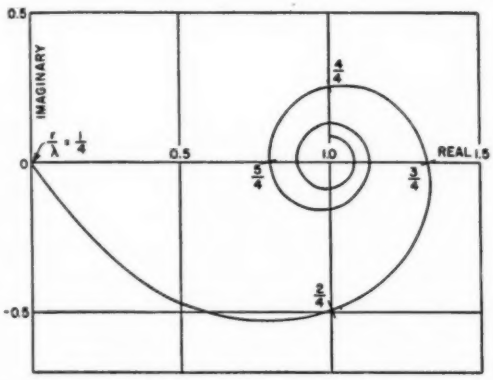


FIG. 4. Spiral for determination of amplitude and phase along a line perpendicular to the direction of incidence and to the axis of the rod at its center. Values of r/λ are indicated by transverse lines along the curve.

e_0 is the amplitude of the incident wave, r_2 and r_3 the distances from either end of the two rods to point P under consideration, t is the time, and T the period. α and k are phase and amplitude constants to be determined by the boundary conditions.

If E_0 and E are the complex rms values of the incident field and the field in the diffraction pattern then

$$E = E_0 + k\left\{\frac{\lambda}{r_2}\exp\left[j2\pi\left(\frac{r_2}{\lambda}+\alpha\right)\right] + \frac{\lambda}{r_3}\exp\left[j2\pi\left(\frac{r_3}{\lambda}+\alpha\right)\right]\right\}. \tag{10}$$

The constants k and α may be evaluated at point A the middle surface of one of the rods where the resultant of the three fields is zero. At A $r_2/\lambda = 1/4$, $r_3/\lambda = [(S/\lambda)^2+(1/4)^2]^{\frac{1}{2}}$. S is the distance between rods.

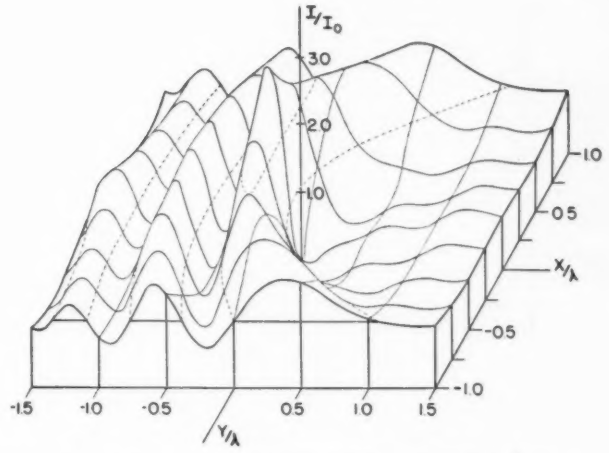


FIG. 5. Experimentally recorded intensities in a plane perpendicular to the rod through its center.

This special value of r_3 will be called r_0. The boundary condition yields $\alpha = 1/4$ and

$$k = \frac{E_0}{4 - \frac{\lambda}{r_0}\exp\left[j2\pi\left(\frac{r_0}{\lambda}+\frac{1}{4}\right)\right]}.$$

Since k is complex, the evaluation of Eq. (10) is time-consuming even for a few special cases. The phases and amplitudes have been expressed by Figs. 7a, b, and c for $S/\lambda = 1$, 2, and 3. Values of d/λ have been indicated along the curve from the boundary at A to the midpoint between the rods. It is to be noted that no attention is given to the finite diameter of the rods. In experiment it was observed that the patterns were the same for rods varying from 0.02λ to 0.06λ in diameter. The small amplitudes at these short distances from the rod indicated in Fig. 7 reveal why the finite size of the rod could be neglected.

Figure 8 is of the calculated intensities for $S/\lambda = 1$, 2, and 3. The experimental data are indicated by x's. These data confirm those of Trantor.[4] Trantor has suggested that the study of diffraction pattern of two rods might indicate an explanation of the diffraction

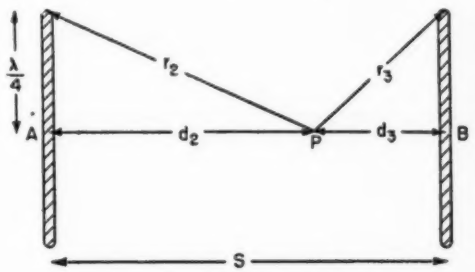


FIG. 6. Cross section through axes of rods.

[4] W. O. Trantor, Phys. Rev. **73**, 184 (1948).

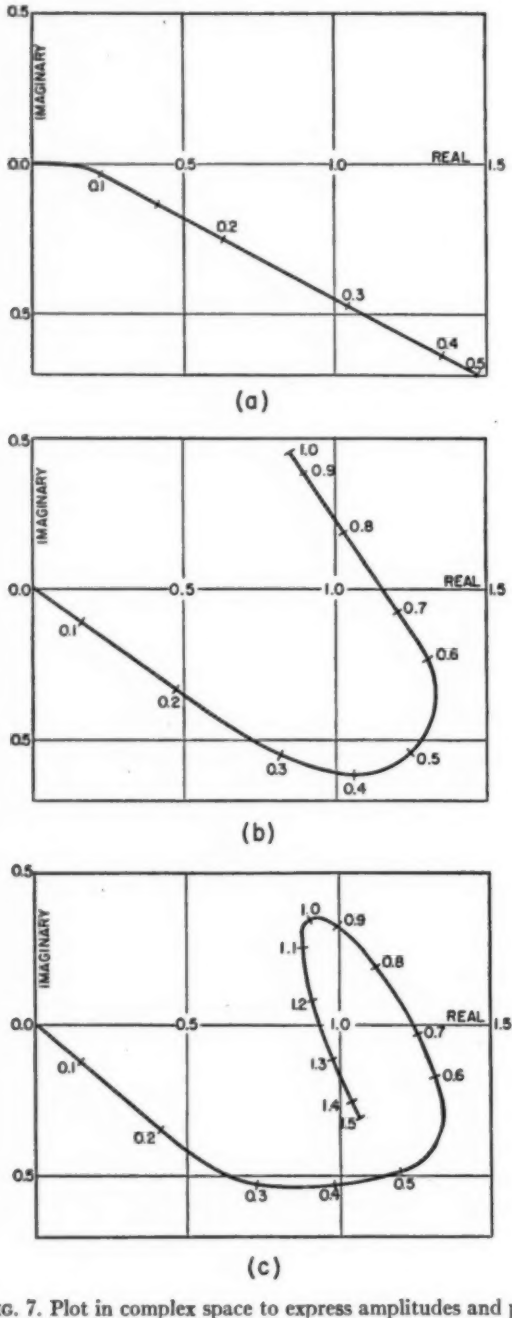


FIG. 7. Plot in complex space to express amplitudes and phases at points along the line between the centers of two rods taken from the boundary A to the midpoint between the rods. a, b, and c are for rods separated by one, two, and three wavelengths.

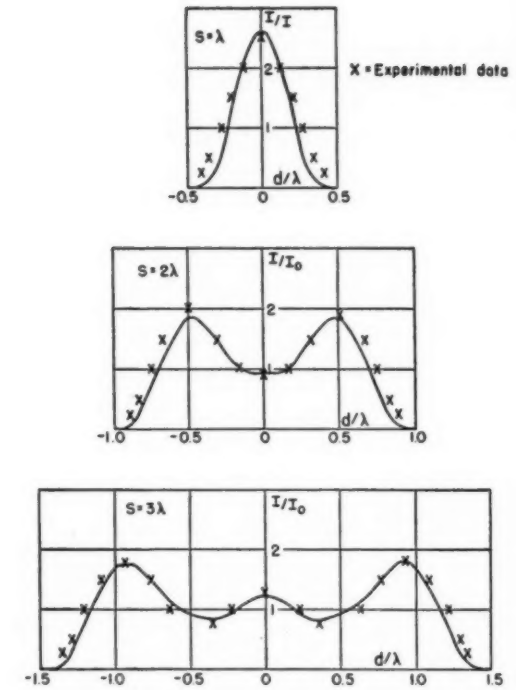


FIG. 8. Comparison of calculated and measured intensities along lines between the centers of the two rods for rods separated by one, two, and three wavelengths.

pattern of a circular aperture. Indeed the positions of maximum and minimum intensity in the magnetic plane through the center of a circular aperture upon which a plane polarized wave is incident normally lie in closely the same positions as the maxima and minima in the diffraction patterns between the pairs of rods.

CONCLUSION

The measurements of the diffraction patterns of microwaves in the neighborhood of rods agree sufficiently with rigorous derivation from electromagnetic theory to suggest that simple microwave measurements will be a fruitful approach to problems of diffraction of electromagnetic waves.

ACKNOWLEDGMENT

It is a pleasure to acknowledge the support throughout this study of W. C. White and E. D. McArthur of the General Electric Research Laboratory.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 4 APRIL, 1951

On the Spectrum of Isotropic Temperature Fluctuations in an Isotropic Turbulence

STANLEY CORRSIN
Department of Aeronautics, The Johns Hopkins University, Baltimore, Maryland
(Received October 6, 1950)

The one- and three-dimensional spectral equations for a field of isotropic temperature fluctuations in an isotropic turbulence are derived from the correlation equation. Then, except for the degenerate case of simple conductive decay, stationary fluctuation fields are postulated. Largely by dimensional reasoning and some simple postulates, forms of the temperature fluctuation power spectrum for particular wave number ranges are deduced and are compared with the results of previous analyses giving corresponding segments of the velocity spectrum. Finally, the relative "effective cut-off" wave numbers of the two spectra are compared in terms of the fluid Prandtl number.

INTRODUCTION

SINCE there is a great deal of interest in the heat and mass transport properties of turbulent fluid flow, an attempt has been made to treat isotropic scalar fluctuations (such as temperature or concentration) in an isotropic turbulence by the methods which have proven fruitful in studying the turbulence itself. Of course, there is no net transport in an isotropic (and homogeneous) field; but statistical analysis of such a case, beside yielding some results of immediate interest, may be by way of introduction to an attack on the considerably more complex problem. At the present time, there is apparently no fundamental (i.e., nonphenomenological) theory of turbulent transport.

The first step in the present program was an analysis of the decay of isotropic scalar fluctuations in a decaying isotropic turbulence.[1] For this the techniques of Taylor,[2] von Kármán and Howarth,[3] and Loitsiansky[4] were used. The present paper is an attempt to deduce the form of the power spectrum (chiefly) of stationary isotropic scalar fluctuations in a stationary isotropic turbulence.

The statistical theory of stationary or quasi-stationary isotropic turbulence in an incompressible continuum has had considerable development during the past decade, beginning with the work of Kolmogoroff,[5] Onsager,[6] and von Weizsäcker,[7] and continuing with that of Heisenberg,[8] Lin,[9] von Kármán,[10] Batchelor,[11] Townsend,[11] Chandrasekhar,[12] and others. Most of this work has been concerned with the prediction of the shape of the power spectrum of the velocity fluctuations in such a random fluid motion. The intractability of the spectral differential equation, including the penalty for the statistical approach (i.e., more unknowns than equations), has necessitated a rather informal and piecemeal treatment. Thus, references 5, 6, and 7 deduced the spectral law in the "intermediate" range of wave numbers, the range for which the incoming and outgoing kinetic energies in the frequency space are equal. Random kinetic energy is fed into the turbulence at wave numbers less than this range and is dissipated to heat by viscous work at wave numbers higher than this range. The $-5/3$-power variation with wave number of both one-dimensional and three-dimensional power spectra follows from dimensional reasoning after the postulate that the spectral form in this range depends only upon the total rate of flow of energy out through it in the frequency space, i.e., the total viscous dissipation in the turbulent field.

When the lowest wave number range is isotropic, a power series development about zero wave number shows that the three-dimensional power spectrum begins with a $+4$-power variation. As pointed out by Lin,[9] Batchelor,[11] and von Kármán,[10] this result is identifiable with the apparent invariance during decay of the product of turbulent energy and fourth moment of the double velocity correlation coefficient.[4]

In a high wave number range where the rate of viscous dissipation to heat dominates rate of transfer of mechanical energy, Heisenberg[8] found a -7-power variation of the spectrum with wave number; and Chandrasekhar[12] has generalized this treatment to include the lower wave numbers. Kovasznay[13] has given a treatment of the smallest eddies which leads to a definite "cut-off" wave number on the high side.

As in reference 1, it will be assumed that the temperature fluctuations are so small as to have no appreciable effect on the velocity field. This is equivalent to postulating an ideal liquid, with equation of state $\rho =$ constant

[1] S. Corrsin, J. Aeronaut. Sci. (to be published).
[2] G. I. Taylor, Proc. Roy. Soc. (London) (A) **151**, 421–78 (1935).
[3] Th. v. Kármán and L. Howarth, Proc. Roy. Soc. (London) (A) **164**, 192–215 (1938).
[4] L. G. Loitsiansky, C.A.H.I. Inst. Moscow, Rept. No. 440 (1939) (Translated as N.A.C.A. T.M. 1079).
[5] A. N. Kolmogoroff, Compt. rend. (Doklady) de l'Acad. Sci. URSS **30**, 301–305 (1941); **32**, 16–18 (1941).
[6] L. Onsager, Phys. Rev. **68**, 286 (1945).
[7] C. F. v. Weizsäcker, Z. Physik **124**, 614 (1948).
[8] W. Heisenberg, Z. Physik **124**, 628 (1948); Proc. Roy. Soc. (London) (A) **195**, 402 (1948).
[9] C. C. Lin, Proceedings of the First Symposium in Applied Mathematics, 1947, published by Am. Math. Soc. (1949); Proc. 7th Inter. Cong. for Appl. Mech. (1948); Th. v. Kármán and C. C. Lin, Revs. Modern Phys. **21**, 516 (1949).
[10] Th. v. Kármán, Compt. rend. **226**, 2108 (1948); Proc. Nat. Acad. Sci. U. S. **34**, 530 (1948).
[11] G. K. Batchelor, Proc. Roy. Soc. (London) (A) **195**, 513 (1949); G. K. Batchelor and A. A. Townsend, Proc. Roy. Soc. (London) (A) **193**, 539 (1948); **194**, 527 (1948); **199**, 238 (1949); and other papers.
[12] S. Chandrasekhar, Phys. Rev. **75**, 896 (1949); Proc. Roy. Soc. (London) (A) **200**, 20 (1949).
[13] L. S. G. Kovasznay, J. Aeronaut. Sci. **15**, 745 (1948).

(ρ is density). Of course, the analysis is equally valid for the concentration fluctuations during isotropic mixing of two equi-dense gases or liquids, provided that the mutual diffusion coefficient is constant and that (for liquids) there is no surface tension between the two ingredients. The treatment here employs temperature fluctuations chiefly for specificity, but also because attempts at experimental verification may be made most easily when the scalar variable is a temperature.

THE SPECTRAL EQUATION

The scalar form of the correlation equation for isotropic temperature fluctuations in an incompresssible isotropic "box" turbulence is[1]

$$\frac{\partial}{\partial t}(m_*)-2\left\{\frac{\partial p_*}{\partial r}+2\frac{p_*}{r}\right\}=2\gamma\left\{\frac{\partial^2 m_*}{\partial r^2}+\frac{2}{r}\frac{\partial m_*}{\partial r}\right\}, \quad (1)$$

where $m_*=\langle\vartheta\vartheta'\rangle_{Av}$, the dimensional correlation between temperature fluctuations at 2 points, P and P'; $p_*=$ the triple correlation between temperature fluctuations at the two points and velocity fluctuation component at P along the PP' direction. This is the characteristic scalar function of the dimensional triple correlation vector, $S_{*i}=\langle u_i\vartheta\vartheta'\rangle_{Av}$; ϑ, $\vartheta'=$ temperature fluctuation at P and P', respectively; $u_i=$ velocity fluctuation vector at P; $t=$ time; $r=$ scalar distance between P and P'; $\gamma=\alpha/\rho c_p$; $\alpha=$ coefficient of thermal conductivity; $\rho=$ density; and $c_p=$ specific heat at constant pressure.

The averaging is a space averaging over the homogeneous field.

An equation for the one-dimensional "energy" (or "power") spectrum of the temperature fluctuation field is most easily obtained by taking the fourier transform of Eq. (1). The fourier transform relationship between autocorrelation and energy spectrum of a stationary one-dimensional random variable was apparently first demonstrated by N. Wiener.[14] It was discovered independently by G. I. Taylor[15] in connection with turbulence theory. We have, for any t,

$$\left.\begin{aligned} m_*(r)&=\int_0^\infty G_1(k_1)\cos(k_1r)dk_1 \\ G_1(k_1)&=(2/\pi)\int_0^\infty m_*(r)\cos(k_1r)dr \end{aligned}\right\}, \quad (2)$$

where $G_1(k_1)=$ the one-dimensional energy spectrum; and $k_1=$ a wave number component in the wave number space (in general, $k=2\pi/l$, where l is wavelength).

Multiplying Eq. (1) by $(2/\pi)\cos(k_1r)$ and integrating over r from 0 to ∞, we wind up with

$$(\partial G_1/\partial t)-2\left\{k_1^2K_1(k_1)-2\int_\infty^{k_1}k_1K_1dk_1\right\} =2\gamma\left\{-k_1^2G_1+2\int_\infty^{k_1}k_1G_1dk_1\right\}. \quad (3)$$

[14] N. Wiener, Acta. Math. **55**, 117 (1930).
[15] G. I. Taylor, Proc. Roy. Soc. (London) (A) **164**, 476 (1938).

The one-dimensional spectral transfer function is defined by

$$\left.\begin{aligned} p_*(r)&=\int_0^\infty k_1K_1(k_1)\sin(k_1r)dk_1 \\ k_1K_1(k_1)&=\frac{2}{\pi}\int_0^\infty p_*(r)\sin(k_1r)dr \end{aligned}\right\}. \quad (4)$$

The equation for $G(k)$, the three-dimensional energy spectrum of the temperature fluctuation field, follows from application of the relation for any t,[16]

$$G(k_1)=-2k_1G_1'(k_1), \quad (5)$$

where the prime signifies differentiation with respect to k_1. The final result is

$$(\partial G/\partial t)-2k^2K=-2\gamma k^2G. \quad (6)$$

The three-dimensional spectral transfer function K is defined by

$$K(k_1)=-2k_1K_1'(k_1). \quad (7)$$

We also note another property of the transfer function, from Eq. (6):

$$\int_0^\infty k^2K(k)dk=0.$$

Equation (6) is of the same form as that for the velocity fluctuation spectrum. Just as in the velocity case, the problem must remain unsolved until an additional relation between energy spectrum and transfer function can be deduced. In the velocity case, various more or less plausible postulates have been made for this required relation,[8,10,13] but no fundamental approach to the problem has yet been possible. Similar postulates can, of course, be made for the present case, but a detailed pursuit of the necessary computations is rather more ambitious than the objectives of this paper.

SOLUTION FOR SMALL PÉCLET NUMBER

A complete solution of the spectral equation can be given only in the degenerate case of very small Péclet number. Péclet number is a measure of the ratio of forced convection effects to conduction effects in a heat transfer problem. A suitable definition in this case is[1] $P_\lambda=[\langle u^2\rangle_{Av}]^{\frac{1}{2}}\cdot(\lambda_\vartheta/\gamma)$, where λ_ϑ is the "microscale" of the temperature fluctuation field. In terms of the one-dimensional power spectrum,

$$1/\lambda_\vartheta^2=(1/\langle\vartheta^2\rangle_{Av})\int_0^\infty k_1^2G_1dk_1. \quad (8)$$

For sufficiently small Péclet number, the transfer term in Eq. (6) is negligible, and Eq. (6) reduces to

$$\partial G/\partial t=-2\gamma k^2G, \quad (9)$$

[16] Kovasznay, Uberoi, and Corrsin, Phys. Rev. **76**, 1263 (1949).

which, of course, just describes the conductive "smearing" of an isotropic temperature "spottiness" in a fixed continuum.

The solution corresponding to that chosen for the velocity field problem at very small Reynolds number[3,4] is

$$G(k,\tau)=(4/\pi)Nk^2\exp(-2\gamma\tau k^2). \tag{10}$$

The three-dimensional energy spectrum of the velocity field at very small Reynolds number (~negligible inertia forces) is, for comparison,

$$F(k,t)=(4/3\pi)Mk^4\exp(-2\nu t k^2), \tag{11}$$

where ν is kinematic viscosity.

In Eqs. (10) and (11), M and N are constants independent of time, evaluated from the Loitsiansky invariant[4] and its thermal analog:[1]

$$M=\langle u^2\rangle_{Av}\int_0^\infty r^4 f(r)dr;\quad N=\langle\vartheta^2\rangle_{Av}\int_0^\infty r^2 m(r)dr. \tag{12}$$

Here, f is the scalar correlation coefficient between velocity fluctuations at two points for components in the direction of the line connecting the points,[3] and $m_*=\langle\vartheta^2\rangle_{Av}m$.

τ and t are both time; different symbols are used to emphasize the fact that even in coexisting velocity and temperature fluctuation fields, the quasi-origins in time of the two random variables will ordinarily not coincide ($\tau=t-t_1$). Both Eq. (10) and Eq. (11) can, of course, be obtained simply from the solutions of the corresponding correlation equations. The relative natures of F and G are shown graphically in Fig. 1 for three different values of the fluid Prandtl number, $\sigma=\mu c_p/\alpha$, where μ=viscosity coefficient. For these curves it is assumed that $t\gg t_1$ so that $\tau\approx t$.

The balance of the paper is devoted to discussion of the spectral form in specific wave number ranges when the convective effects are not necessarily negligible.

LOWEST WAVE NUMBER RANGE

Since for any Péclet number, the correlation function $m_*(r)$ is an even function, its fourier cosine transform, $G_1(k_1)$, is also an even function. A power series expansion about $k_1=0$ has the form

$$G_1(k_1)=G_1(0)+G_1''(0)\frac{k_1^2}{2!}+G_1^{IV}(0)\frac{k_1^4}{4!}+\cdots. \tag{13}$$

Just as in the case of the velocity field, $G_1(0)$ is proportional to the "scale" of the temperature fluctuation field. The scale is defined by

$$\Lambda(t)=(1/\langle\vartheta^2\rangle_{Av})\int_0^\infty m_*dr. \tag{14}$$

Letting $k_1\to 0$ in the second equation of Eq. (2) and assuming a suitably rapid decrease to zero of $m_*(r)$ at large r, we find

$$G_1(0)=(2/\pi)\langle\vartheta^2\rangle_{Av}\Lambda \tag{15}$$

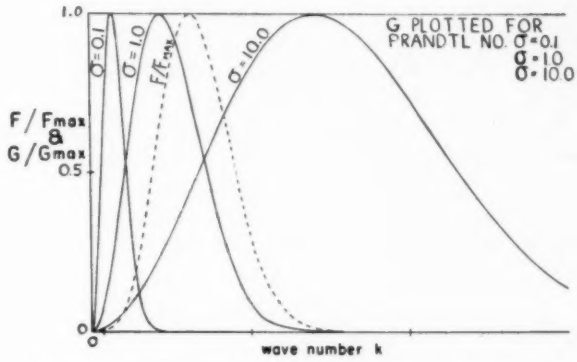


FIG. 1. Temperature and velocity spectra for very small Reynolds and Péclet numbers.

similar to the well-known expression for the velocity scale.

Application of Eq. (5) to Eq. (13) gives the k-behavior of the three-dimensional energy spectrum around the origin at any time:

$$G(k_1)=-2G_1''(0)k_1^2-\tfrac{1}{3}G_1^{IV}(0)k_1^4-\cdots;$$

therefore,

$$G(k)=-G''(0)\frac{k^2}{2!}-G^{IV}(0)\frac{k^4}{4!}-\cdots. \tag{16}$$

Thus, the temperature spectrum begins parabolically, in contrast to the fourth-power start of the three-dimensional energy spectrum of the isotropic velocity fluctuations.[9,11] This parabolic behavior is seen to be consistent with the nature of the invariant N in Eq. (12), just as the quartic behavior of the velocity spectrum is consistent with the Loitsiansky invariant M in Eq. (12). We note also that the spectral forms Eqs. (10) and (11) for small Péclet and Reynolds numbers, respectively, demonstrate the proper behavior as $k\to 0$.

THE INTERMEDIATE WAVE NUMBER RANGE

For sufficiently high Péclet numbers of the temperature and velocity fluctuation field, we postulate the existence of an "intermediate" wave number range which is high compared with the largest fluctuations present and low compared with the smallest fluctuations present. Conceptually, this is analogous to the intermediate range in the dynamical problem, although not coincident with it except when the fluid Prandtl number is unity.

All of the fluctuation "energy" entering such a range passes through its lower bound from the lower wave numbers, and it all goes on to the higher wave numbers with no appreciable "dissipation" caused by molecular effects. Since we are examining temperature fluctuations, the mean square of our physical variable $\langle\vartheta^2\rangle_{Av}$ is not really energy in the sense that $\langle u^2\rangle_{Av}$ is energy, and the molecular action (smearing out of fluctuations because of thermal conductivity) is not dissipation in the literal sense. In fact, it is perhaps not intuitively

obvious that the temperature fluctuation "energy" is conserved as it migrates outward through the wave number space. However, the existence of conservation Eq. (6) shows that this is indeed the case.

For the intermediate wave number range, we postulate that the spectrum $G(k)$ depends only upon

(a) the "energy" flow rate through this range in the wave number space, which is simply equal to the total rate of "dissipation" of temperature fluctuations,

$$D_\vartheta = -2\gamma \int_0^\infty k^2 G dk; \quad \text{and}$$

(b) the energy flow rate through the "intermediate" range of the velocity spectrum, which is simply equal to the total rate of dissipation of velocity fluctuations,

$$D = -2\nu \int_0^\infty k^2 F dk.$$

The second postulate renders this analysis valid only in the spectral region where the two "intermediate" wave number ranges overlap.

The dimension of the pertinent quantities are

$$\left.\begin{aligned} |k| &= L^{-1} \\ |G| &= L\Theta^2 \\ |D_\vartheta| &= \Theta^2 T^{-1} \\ |D| &= L^2 T^{-3} \end{aligned}\right\}, \quad \text{where} \quad \begin{cases} L \equiv \text{length} \\ \Theta \equiv \text{temperature} \\ T \equiv \text{time}. \end{cases}$$

Hence, the only possible arrangement is

$$G(k) = A D_\vartheta D^{-\frac{1}{2}} k^{-5/3}, \tag{17}$$

where A is a dimensionless constant.

The corresponding result for the velocity field, assuming that $F(k)$ depends only upon D, gives the well-known result,

$$F(k) = B D^{\frac{1}{2}} k^{-5/3}, \tag{18}$$

where B is a dimensionless constant.

Equation (17) follows also with an alternative postulation, i.e., if (b) is replaced by dependence upon F and use is made of Eq. (18).

Rough estimates can be made of constants A and B through replacing the complete spectrum by an "equivalent" intermediate type ($\sim k^{-5/3}$) spectrum.* This leads to estimates of order unity for both.

HIGH WAVE NUMBER RANGE

In the range of wave numbers where molecular conduction effects are not negligible compared with forced convection effects, we can set up for the temperature fluctuations an equation similar to that of Heisenberg[8] for the corresponding region in the velocity spectrum.

The rate of transfer of temperature fluctuation "energy" through wave number k is thus

$$\epsilon_{\vartheta k} = 2\{\gamma + \gamma_k\} \int_0^k G(k) k^2 dk. \tag{19}$$

From dimensional reasoning, identical with that of Heisenberg for the velocity field, we write

$$\gamma_k = \kappa_\vartheta \int_k^\infty [F(k)/k^3]^{\frac{1}{2}} dk, \tag{20}$$

where the turbulent spectral transfer function κ_ϑ is assumed to be a constant.

Substituting Eq. (20) into Eq. (19), we differentiate the result with respect to k, remembering that $\epsilon_{\vartheta k}$ is independent of k for wave numbers large compared with the smallest ones present in the spectrum. Then,

$$\frac{\gamma}{\kappa_\vartheta} + \int_k^\infty \left[\frac{F}{k^3}\right]^{\frac{1}{2}} dk = \frac{F^{\frac{1}{2}}}{G k^{7/2}} \int_0^k k^2 G dk. \tag{21}$$

This should be compared with Heisenberg's equation for F, written in identical form:

$$\frac{\nu}{\kappa} + \int_k^\infty \left[\frac{F}{k^3}\right]^{\frac{1}{2}} dk = \frac{F^{\frac{1}{2}}}{F k^{7/2}} \int_0^k k^2 F dk, \tag{22}$$

and κ is the (assumed constant) transfer function for F.

Obviously, $G(k)$ in Eq. (21) has a solution similar to that for $F(k)$ obtained from Eq. (22) when the latter is substituted into Eq. (21). A complete solution of Eq. (22) has been given by Chandrasekhar,[12] showing the expected behavior like $k^{-5/3}$ when viscous effects are negligible and Heisenberg's k^{-7} solution when viscous effects dominate. Since G has identical behavior, it is clear that this approach yields

$$G(k) \sim k^{-7} \tag{23}$$

in the range where molecular conduction dominates and where $F \sim k^{-7}$.

It may be remarked at this point that Kolmogoroff's postulate of "local isotropy" in a non-isotropic turbulence[5] is doubtless equally valid for non-isotropic temperature fluctuations in such a turbulence.

THE "CUT-OFF" WAVE NUMBERS

Although there is no obvious reason to suppose that the power spectra literally cut off at any specific wave number on the high side, some insight can be gained into the relative "fine structures" of the velocity and temperature fields by examination of the spectra in greatly simplified form. We replace the actual spectrum by an "equivalent" intermediate type spectrum. Alternatively, this may be regarded as a field of such high Reynolds and Péclet numbers that the $-5/3$ region extends over most of the spectrum. As long as $k_c \gg k_0$, the exact value of k_0 is unimportant here (Fig. 2). The

*I first encountered this simplified approach in a lecture on turbulence by Dr. J. von Neumann, at Princeton University, October, 1949.

existence of lower critical Reynolds numbers for laminar flows certainly suggests that beyond some wave numbers in the velocity spectrum virtually all energy will be dissipated directly to heat, without first breaking up into kinetic energy of still smaller eddies. This lends some physical feeling to ones concept of an effective cut-off wave number for F. It could also be given a specific definition in terms of geometry of the spectral curve. However, for our purposes, the cutoff concept may simply be associated with the fact that at sufficiently high wave numbers the (spectrally) local phenomena are no longer dominated by inertia forces; the viscous effects become important. We shall arbitrarily choose cutoff as the wave number at which the local spectral Reynolds number is 10. As a matter of fact, this appears to be the order of magnitude of the lower critical Reynolds number for stability of laminar flows without solid boundary.[17]

The possible basis of an effective cutoff in the temperature spectrum is a rather less orthodox question. It is physically clear that with Prandtl number σ different from unity the effective cutoffs should correspond to different wave numbers: for very large conductivity ($\sigma \to 0$), the temperature spectrum must cut off at much lower wave number than the velocity spectrum; for very small conductivity ($\sigma \to \infty$), the temperature spectrum will be governed by forced convection effects (as opposed to conduction effects) to indefinitely high wave numbers, and, thus, will "cut off" at much higher wave numbers than the velocity spectrum. The $\sigma = \infty$ case is simply the turbulent "intermingling" (=mixing) of two molecularly immiscible liquids with zero surface tension between them.

Considering the effective cutoff in temperature spectrum as a measure of where conductive effects become important, we base the definition on a value of local (spectral) Péclet number. Following the particular Reynolds number choice for the velocity field, we arbitrarily assume the effective cutoff where the local Péclet number is 10.

Dimensional considerations suggest the following definitions of local (spectral) Reynolds and Péclet numbers:

$$R_k = (kF)^{\frac{1}{2}} l/\nu = (2\pi/\nu)(F/k)^{\frac{1}{2}}; \tag{24}$$

$$P_k = (kF)^{\frac{1}{2}} l/\gamma = (2\pi/\gamma)(F/k)^{\frac{1}{2}}. \tag{25}$$

With k_c and ${}_\vartheta k_c$ symbolizing the effective cut-off wave numbers,

$$10 \approx (2\pi/\nu)[F(k_c)/k_c]^{\frac{1}{2}}; \tag{26}$$

$$10 \approx (2\pi/\gamma)[F({}_\vartheta k_c)/{}_\vartheta k_c]^{\frac{1}{2}}. \tag{27}$$

Substituting Eq. (18) into the ratio of Eq. (27) to Eq. (26), we obtain an expression for the ratio of the cut-off wave numbers:

$$ {}_\vartheta k_c/k_c = \sigma^{\frac{1}{2}}. \tag{28}$$

Obviously, this result behaves in the limits ($\sigma \to 0, \infty$) as anticipated.

A qualitative check can be made on this result by using it to estimate the relative microscales of velocity and temperature fields. The microscales are defined in terms of the one-dimensional spectra:

$$\frac{1}{\lambda^2} = \frac{1}{\langle u^2 \rangle_{Av}} \int_0^\infty k_1^2 F_1 dk_1; \quad \frac{1}{\lambda_\vartheta^2} = \frac{1}{\langle \vartheta^2 \rangle_{Av}} \int_0^\infty k_1^2 G_1 dk_1. \tag{29}$$

But, for our highly simplified spectra, with $F(k)$, $G(k) \sim k^{-5/3}$ between lower and upper cut-off wave numbers, both of the one-dimensional spectra in Eq. (29) are proportional to $k_1^{-5/3}$ everywhere that they are non-zero. Hence, for given $\langle u^2 \rangle_{Av}$ and $\langle \vartheta^2 \rangle_{Av}$,

$$\left(\frac{\lambda_\vartheta}{\lambda}\right)^2 \sim \frac{\int_0^{k_c} k_1^{\frac{1}{3}} dk_1}{\int_0^{{}_\vartheta k_c} k_1^{\frac{1}{3}} dk_1} \sim \left(\frac{k_c}{{}_\vartheta k_c}\right)^{4/3}.$$

According to Eq. (28), this gives

$$(\lambda_\vartheta/\lambda)^2 \sim 1/\sigma \tag{30}$$

which is in agreement with the results deduced by independent methods from the correlation equations for the cases of very low Reynolds and Péclet numbers and very high Reynolds and Péclet numbers.[1]

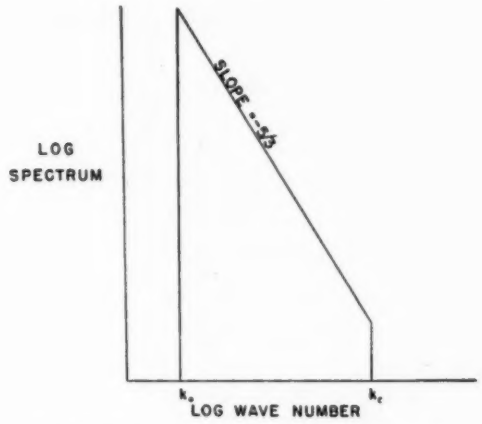


FIG. 2. "Intermediate-type" spectrum $k_c \gg k_0$.

ACKNOWLEDGMENT

I should like to thank Dr. F. H. Clauser and Mr. G. Mélèse for their helpful criticism.

[17] M. Lessen, N.A.C.A., Tech. Note No. 1929 (1949).

Lubrication at Extreme Pressures with Mineral Oil Films

O. H. Clark, W. W. Woods, and J. R. White
Socony-Vacuum Laboratories, Research and Development Department, Paulsboro, New Jersey
(Received October 23, 1950)

Intact mineral oil films at average pressures ranging from 86,000 to 215,000 psi and at high rates of shear have been obtained using a bearing consisting of the contact area between crossed axis steel cylinders. Intact films have also been maintained at low rates of shear and lower pressures using a bearing consisting of a sphere in a spherical shell. The lubricating films possess the shearing properties of plastic solids, a fact which is consistent with solidification observations reported in the literature of high pressure viscometry. In the crossed cylinder bearing the high film pressures required for lubrication are induced by hydrodynamic forces. Such forces can be very large as a result of the rapid increase in viscosity incurred by mineral oils under increasing pressure. The observation of complete mineral oil lubrication at such extreme pressures resulting from known bulk properties, namely, viscosity increase and solidification under pressure, suggests that these properties are sufficient to account for practical mineral oil lubrication.

INTRODUCTION

CONSIDERABLE evidence exists which indicates that mineral oils can, under certain conditions, supply effective lubrication at pressures beyond the range of known hydrodynamic bearing practice. For example, Blok reported[1] that wear of deliberately misaligned and impacting spur gears was profoundly influenced by base oil viscosity although essentially unaffected by addition of fatty acids or a variety of extreme pressure additives. In direct impact experiments Bowden and Tabor have shown[2] that unconfined oil films can transmit pressure sufficient to cause plastic flow of copper without film rupture. A hypothetical mechanism by which mineral oils may provide protection from wear is indicated by the results of high pressure viscometry studies[3–7] which show that these oils undergo very rapid viscosity increase with increasing pressure and that at least some of them solidify at pressures attainable in lubrication processes.

The present paper reports experiments demonstrating lubrication by intact mineral oil films at pressures ranging from 86,000 to 215,000 psi. Evidence is adduced that the lubricating oil films are essentially solid and that the pressures necessary to solidify the oils are created in the bearing by hydrodynamic forces.

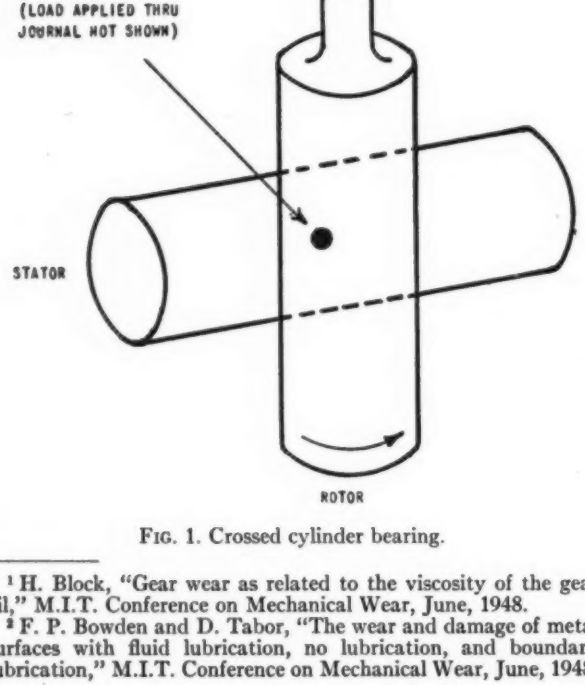


FIG. 1. Crossed cylinder bearing.

PART I. CROSSED CYLINDER BEARING EXPERIMENTS

Apparatus

In the first part are described experiments on a crossed cylinder bearing developed for lubrication studies. A schematic drawing of the bearing used is shown in Fig. 1. This bearing consists of the small contact area between a $\frac{1}{4}$-in.-diameter cylindrical steel rotor and an identically shaped stator oriented with its axis at right angles to that of the rotor. The same geometry was previously used by Burwell and Strang[8] who described the advantages of mating convex surfaces in obtaining a bearing of well-defined effective area which is free from widely scattered point contacts and is amenable to initial surface finish control. The presently described bearing, however, differs from theirs in its use of a hardness differential between rotor (65 Rockwell C) and stator (90 Rockwell B) to preserve an unchanging rotor finish and minimize mutual tearing and metal adhesion. This feature is highly desirable in the present experiments where the objective is to accentuate rather

[1] H. Block, "Gear wear as related to the viscosity of the gear oil," M.I.T. Conference on Mechanical Wear, June, 1948.
[2] F. P. Bowden and D. Tabor, "The wear and damage of metal surfaces with fluid lubrication, no lubrication, and boundary lubrication," M.I.T. Conference on Mechanical Wear, June, 1948.
[3] M. D. Hersey and H. Shore, Mech. Eng. **50**, 221–232 (1928).
[4] R. V. Kleinschmidt, Trans. ASME, **50**, APM 50–4, 2–5 (1928).
[5] R. B. Dow, J. Appl. Phys. **8**, 367–372 (1937).
[6] R. T. Sanderson, Ind. Eng. Chem. **41**, 2059 (1944).
[7] M. D. Hersey and R. F. Hopkins, Mech. Eng. **67**, 820–824 (1945). (This compendium has recently been greatly expanded but as of this date not yet republished.)
[8] J. T. Burwell and C. D. Strang, J. Appl. Phys. **20**, 79–89 (1949).

than—as Burwell and Strang desired—to minimize hydrodynamic effects.

Reference to the general apparatus diagrams, Figs. 2 and 3 will indicate the drive, loading, and torque indicating devices. Two drives have been used, one employing frequency changer ac-motor speed control and the other thymatrol dc-motor control. Rotation is transferred to a secondary shaft mounted in sleeve bearings by means of a vibration isolating rubber coupling and thence imparted to the rotor itself through a light short-coupled double universal joint. Load is transmitted to the rotor through a 180° sleeve bearing, swivel-mounted on a hydraulically actuated nutcracker. Loading bearings are subjected to prior running-in as pairs at loads greatly in excess of test loads so that they will themselves operate hydrodynamically when put into service. At the conclusion of this run-in operation, their friction coefficients are measured over a range of conditions, and the values so obtained are used to correct total friction measured in subsequent crossed cylinder runs. The corrections involved are small, loader coefficients being of the order of 0.003 contrasted to coefficients of around 0.070 for the crossed cylinder bearing itself. Friction measurements are provided for by a construction wherein the stator is rigidly affixed to the test oil cup which itself is mounted on thrust bearings and is free to rotate on an axis coincident with that of the rotor. External restraint of this motion by a Statham wire strain gauge then provides the means of measuring total torque.

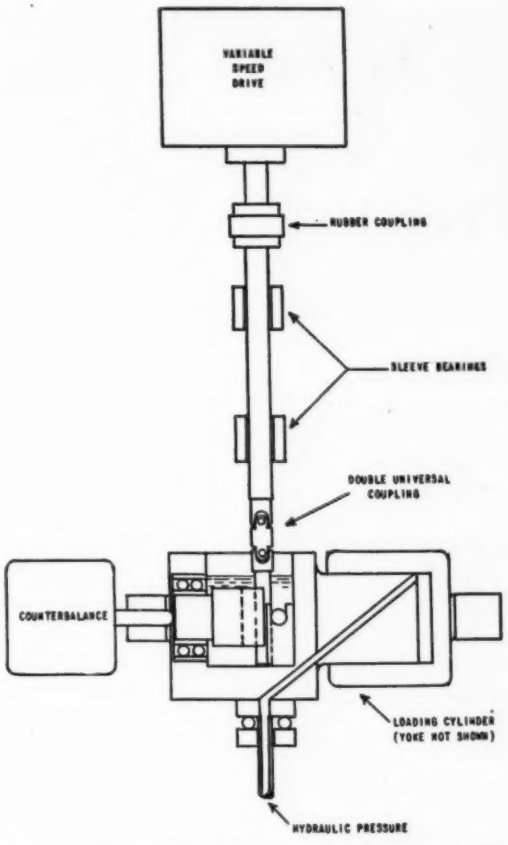


FIG. 2. Crossed cylinder bearing apparatus vertical view.

Operational Features

In order to obtain maximum hydrodynamic support in the crossed cylinder bearing, it is necessary to minimize disturbing torques and thrusts from the loading and driving systems. Since the contact area is so small (usually less than 1/64-in. diameter), torques arising within the bearing are very small and bearing equilibrium can be disturbed by very small external forces having a large mechanical advantage.

For these reasons it was found necessary to guard against eccentricities in rotor motion caused by motor vibration, off-axis drives, and eccentricity of the rotor itself. Drive motors were specially balanced and their rotation transferred to a secondary drive shaft in well-fitted sleeve bearings through a vibration isolating rubber coupling. Stator location was arranged so that when rotor-stator contact occurred the rotor and drive shaft axes coincided within 0.001 in. Since under some operating conditions the stator may be worn to a depth of several ten thousandths of an inch, further precision of alignment was not attempted. To minimize the effect of misalignment in inducing conical rotor movement, a light short-coupled double universal joint was employed to join rotor and secondary drive shaft. Periodic thrusts from rotor out-of-round were largely eliminated by specifying a maximum out-of-round tolerance of 0.00005 in.

An additional requirement for successful operation is that the loading device must be able to exert a thrust of variable direction. This necessity arises because the magnitude of the frictional component of force in the crossed cylinder bearing varies during a run and hence the direction of the resultant force in the bearing must also vary. To meet this requirement with the hydrodynamic loading bearings used, a small amount of lateral displacement proved to be necessary and was provided by elasticity in the loading bearing mounting. A check on the hydrodynamic operation of these loading bearings was obtained by observing the absence of scores on extremely thin copper films plated on their surfaces.

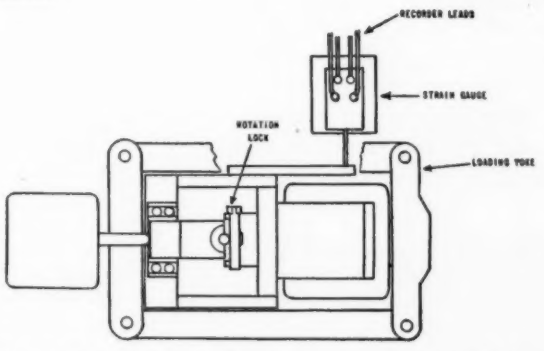


FIG. 3. Crossed cylinder bearing apparatus horizontal view.

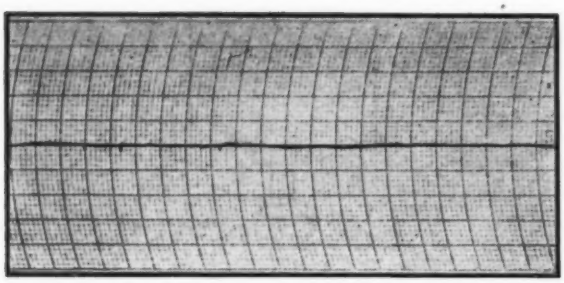


FIG. 4. Brush profilograph of 8μ-diamond rotor finish. One vertical division equals one micro-inch.

Surface finish requirements varied widely with operating conditions. For many experiments it was necessary to polish rotor surfaces with fine diamond abrasives to finishes so smooth that no rms reading could be obtained with commercial profilometers. Figure 4 is the Brush surface analyzer pattern of one such finish, produced in this instance by polishing with an 8μ-abrasive. On occasion one micron and even $\frac{1}{4}\mu$-grades have been used. The smooth rotor finishes produced were preserved virtually intact throughout runs, as evidenced by reproducibility of repeat runs on the same track as well as by microscopic examination. For most purposes surface finish of the stator was not critical because the hard rotor very quickly wore away the stator's original finish, and 4/0 emery finishes were sufficient.

Intact Oil Films at Extreme Pressures

A salient feature of lubrication studies made with the crossed cylinder bearing has been the observation of intact mineral oil films at very high pressures. The existence of such films is demonstrated by the absence or the cessation of mechanical wear. This is graphically illustrated in Fig. 5, which is a photo-micrograph of the contact area (or "pit") resulting from running the bearing 10,000 revolutions at 800 rpm and 23.6-lb load with a Duosol refined midcontinent oil having a viscosity of 480 centistokes at 100°F. The pit diameter is 0.0118 in., corresponding to an average bearing pressure of 215,000 psi; yet, the 4/0 emery finish of the stator is virtually intact.

The aforementioned use of pit diameter to calculate average bearing pressure must be justified by showing that virtually all the load is supported within the visibly deformed area. To this end careful measurements

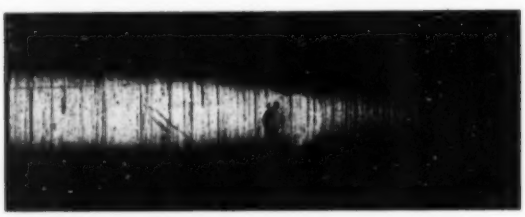


FIG. 5. High pressure mineral oil lubrication without wear. (Magnification 25×; average bearing pressure 215,000 psi.)

were made comparing the diameters of pits from no-wear mineral oil runs with those of pits made by plastic deformation under dry static loading. Measured values are given in Table I, and show that the residual pits ensuing from plastic flow are the same size in both cases. It follows that the load fraction supported within the visible residual pit area while running on an oil film is the same as in the case of direct metallic contact. In the latter case, loads this far in excess of the elastic limit are almost entirely supported on the area of plastic flow. Under load conditions typical of crossed cylinder bearing operation, Clark and Lyman[9] found that the residual pit area was 90 percent of the contact area under load. Moreover, the pressure distribution over the contact area is not uniform but is a maximum in the center. The contact contour for the closely related geometry of a ball upon a plate has been investigated[10] and the results indicate a hemispherical pressure distribution.[11] Assuming such a pressure distribution in this case, an estimate was made of the load fraction supported in the visible pit. The resulting fraction is 97 percent. To a good approximation, therefore, bearing pressure may be taken as the ratio of load to residual pit area.

TABLE I. Static *vs* dynamic pit diameters.

Load	Static pit diameter	Dynamic pit diameter
15.7 lb	0.0257 cm	0.0254, 0.0255, 0.0258, 0.0258, 0.0250, 0.0255, 0.0253, 0.0253, 0.0260 cm
40.25	0.0386	0.0384, 0.0379, 0.0391

TABLE II. Wear cessation.

Revolutions	Pit diameter	Average pressure	Wear increment[a] per revolution
0	0.0257 cm	195,000 psi	...
4,500	0.0366	96,000	264×10^{-9} mg
9,000	0.0378	90,300	52.5×10^{-9}
27,000	0.0378	90,300	0×10^{-9}
81,000	0.0386	86,400	2.8×10^{-9}

[a] Calculated from the following relation expressing the common volume between two intersecting cylinders: $V=(3/256)\pi d^4/a$ where d = pit diameter and a = cylinder radius. The effect of elastic deformation has been neglected.

Under many sets of conditions the plastically formed contact area is insufficient for intact oil film operation and mechanical wear ensues. The pit is then enlarged until bearing pressures are sufficiently low for the re-establishment of intact film lubrication. This cessation of wear can be demonstrated by measuring pit diameters for a series of runs of successively increased duration. A typical set of results is given in Table II. These data were obtained at 1800 rpm and at a 15.7-lb load using a

[9] O. H. Clark and E. M. Lyman, J. Appl. Phys. **20**, 884-885 (1949).

[10] D. Tabor, Proc. Roy. Soc. (London) **192A**, No. A1029, p. 247 (1948).

[11] S. Timoshenko, *Theory of Elasticity* (McGraw-Hill Book Company, Inc., New York, 1934).

midcontinent Duosol refined residuum of 480 cs viscosity at 100°F and a 2/0 emery rotor surface finish.

Torque measurements taken throughout runs indicate a corresponding decline in friction during the pit growth period followed by stabilization when wear has been completed. An example of torque recession is given in Fig. 6. This particular run represents bearing operation with a midcontinent Duosol refined residuum (222 cs at 100°F) at 50 rpm, a load of 23.6-lb and an 8μ-rotor finish.

A photo-micrograph of a typical pit after wear has ceased is shown in Fig. 7. Removal of metal was very nearly confined to erasure of the 4/0 emery finish, as exhibited in the photo-micrograph; and its amount calculated by the formula in the previous footnote is about 0.0003 mg. This run was made under conditions identical to those used to produce the uncut pit of Fig. 5 except for the substitution of a similar lubricant of lower viscosity (65 cs at 100°F). The resulting contact area corresponds to an average bearing pressure of 178,000 psi.

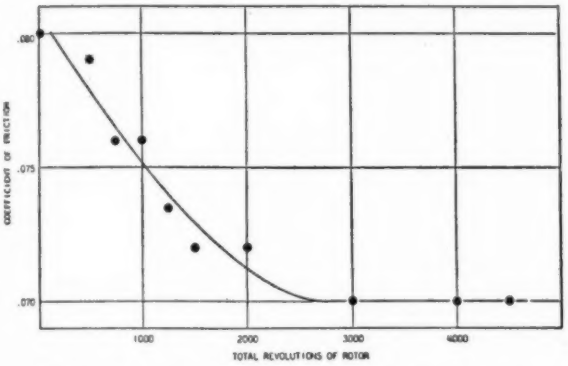


FIG. 6. Torque recession in crossed cylinder bearing.

The Mechanics of Lubrication by Mineral Oils

The mechanism by which mineral oils lubricate in the crossed cylinder bearing is not obvious, since as Burwell and Strang[8] pointed out, the geometry of this bearing is particularly unfavorable to hydrodynamic load carriage. Despite this unfavorable geometry, however, the primary mechanism for the formation of these oil films, which fully separate the bearing surfaces, as illustrated in Fig. 5, is hydrodynamic. Evidence supporting this conclusion is presented below.

The relation between lubrication and bearing speed is of particular importance because a decrease in bearing speed, other variables remaining unchanged, can lead to increased wear only when a hydrodynamic mechanism operates. Measurements were made of pit diameters after wear cessation for wide ranges of speed under otherwise identical operating conditions. For every oil tested equilibrium pit diameters decreased systematically with increased speed up to a critical speed beyond which further decreases in pit diameter did not occur. A lower limit on pit diameter is obviously imposed by plastic flow of the bearing metal, regardless of operating conditions. Figure 8 illustrates this behavior for operation at 23.6-lb load and an 8μ-diamond rotor surface finish with a midcontinent Duosol refined residuum of 222 cs viscosity at 100°F. In these particular experiments the minimum pit diameter attained is 0.0327 cm. Plastic flow of the stator under the given load establishes a slightly smaller lower limit of 0.0310 cm. Except under the most favorable operating conditions, plateau pit diameters exhibited similar small excesses over the static plastic pit size. These excesses apparently arise from surface abrasion occurring during the short interval of bearing formation by plastic flow. If a highly viscous oil is used during this formative period and is then completely replaced while running with the less viscous test oil by means of successive transfusions, the resulting pit diameters do indeed approach the limiting plasticity values.

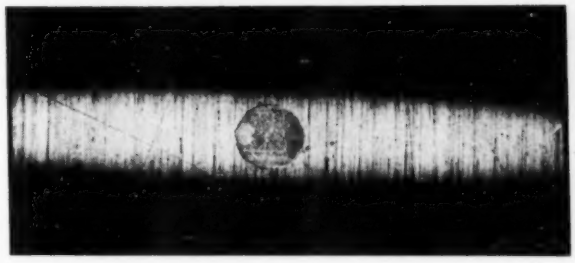


FIG. 7. Equilibrium pit formed by mechanical wear. (Magnification 25×; average bearing pressure 178,000 psi.)

In experiments on the effect of speed variation oil film temperatures were not identical for the various speeds. Operating practice was to keep bulk oil temperature constant, and hence the greater heat input at

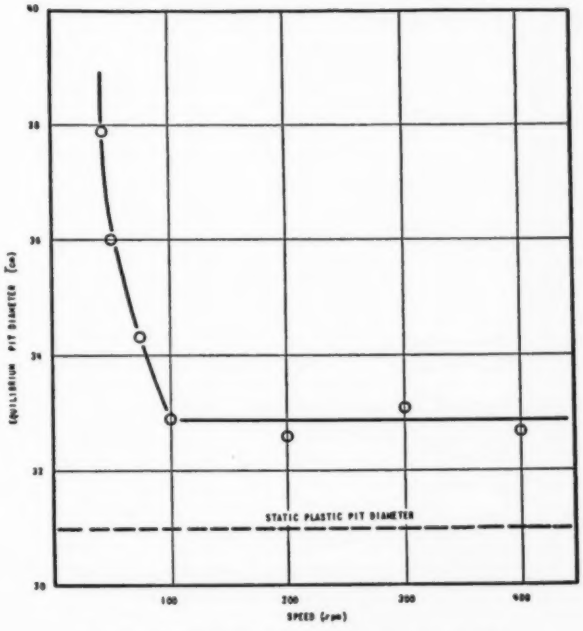


FIG. 8. Equilibrium pit sizes for various speeds.

TABLE III. Viscosity and wear.*

Viscosity cs at 100°F	Equilibrium pit size	Equilibrium pressure
222.5	0.0327 cm	181,000 psi
69.8	0.0361	149,000
20.2	0.0431	105,000

* Operating conditions, 23.6-lb load, 300 rpm, 8μ-diamond finish.

TABLE IV. Properties of oils for high pressure viscosity tests.

Oil	Viscosity cs at 100°F	Viscosity cs at 210°F	VI
A	222.5	17.92	95.4
B	217.6	12.67	23.4

higher speeds led to somewhat greater film temperatures. This film temperature increase could only detract from the favorable influence of speed upon preserving intact oil films since it would lead to relative viscosity loss at higher speeds. Inasmuch as a marked positive speed response in reducing stator wear was observed, no efforts were made to adjust to equal oil film temperatures.

In hydrodynamic mechanisms viscosity increase can augment load carriage in the same manner as does speed increase. Table III compares equilibrium pit sizes for oils of widely different viscosities and clearly establishes that this effect exists in the crossed cylinder bearing.

A rigorous study of viscosity effects would require knowledge of viscosity at the very high pressures existing in the bearing. Such knowledge is not available. However, published data[3–7] show that in general there is approximately an exponential increase in viscosity with increasing pressure and that the rate of viscosity-pressure response usually correlates with the degree of viscosity-temperature response. The existence of such a correlation under lubricating conditions was investigated by comparing the crossed cylinder bearing performance of two oils possessing nearly equal room-pressure viscosities at test temperatures but different viscosity indices. Physical characteristics of the oils used are listed in Table IV.

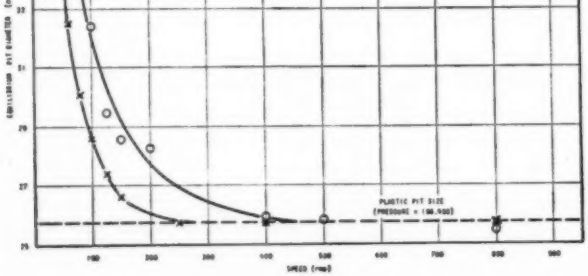


FIG. 9. Effect of high pressure viscosity on bearing performance.

These two oils were compared by making a series of measurements of equilibrium pit diameters for various operating speeds. Results shown in Fig. 9 indicate that the oil with the lower viscosity index was the more effective lubricant, although its room-pressure viscosity is slightly less at the operating temperatures. At 450 rpm, the speed at which lubrication performances began to diverge, calculated* film temperature for both oils was 116.6°F. Room pressure viscosities at this temperature would be 155 cs for oil A and 140 cs for oil B. Nevertheless, oil B was a more effective lubricant under these conditions. At 275 rpm, the speed at which wear began with oil B, calculated film temperature was 102.3°F and the corresponding room pressure viscosity 200 cs Oil A at this temperature again has slightly higher room-pressure viscosity (205 cs) than has B, but again A was the less effective lubricant.

Film Thickness, Temperature, and Rate of Shear in the Crossed Cylinder Bearing

In order to arrive at an estimate of equilibrium clearance in the crossed cylinder bearing, a hypothetical picture of its operation has been adopted which is closely analogous to the operation of a bedded brass bearing. It is first noted that so long as wear is occurring over the entire stator pit the clearance is effectively limited by the roughness of the rotor, continual contact being made by the highest asperities on the rotor. It is assumed that at some time during this cutting period gradual formation of an oil wedge starts. At first, while the lift from the forming oil wedge is negligible, cutting will continue over the entire pit; but as soon as the oil wedge starts to carry appreciable load, cutting will become preferential toward the exit edge of the bearing, and rotor displacement will result. Further enlargement of the bearing will consist principally of cutting at this edge, and therefore of elongation of the pit in the direction of rotation. The process of oil wedge development is concurrent with rotor displacement and its accompanying preferential cutting of the exit edge. With this picture the amount of rotor displacement and the corresponding clearance could be determined by measurement of the eccentricity of the pit.

Two methods were employed in an effort to detect this eccentricity. The first consisted of direct comparison of pit diameters parallel and transverse to the direction of rotation. Although there were random small variations due to stator and rotor imperfections, no trend toward unequal diameters was found. Within measuring precision, about 0.001 cm, the pits were round.

The second method consisted of mapping the bearing contours by interference fringes. Details of this procedure were described by Clark and Lyman.[9] Sample photographs are shown in Fig. 10. From such contour maps, pit depths were obtained and compared with the depths of the rotor arcs whose chords were set equal to

* The method of calculation is given in the following section.

the measured pit diameters. Results are given in Table V. Within the precision of the method, pit depths match rotor depth requirements. The precision of depth determination by contour mapping is about $\pm 3\times 10^{-5}$ cm. The uncertainty in diameter resulting from this uncertainty in depth varies inversely as the diameter itself. By way of example, for a diameter of 0.034 cm, this uncertainty is about ± 0.0011 cm.

Thus by both methods, direct measurement and pit contour mapping, rotor displacement evidenced by preferential cutting could not be detected. The limit of measuring precision was therefore taken as an upper limit for estimating bearing clearance. For a typical pit of 0.034 cm the clearance at the center of the pit corresponding to rotor displacement of 0.001 cm is roughly 5×10^{-5} cm, circular pit contour having been assumed for simplicity in calculation. This figure is believed to be a fair order of magnitude upper limit for bearing clearance after wear cessation.

Temperatures of oil films in the crossed cylinder bearing have been calculated on the assumption that the frictional heat input is generated uniformly over the pit area. The oil film is then treated as an infinitely thin circular heat source in an infinite mass of conducting metal. It is implicit in this treatment that heat generated in the oil film is conducted into the metal too rapidly to be carried away with the oil. Since the film thickness is of the order of a thousand times smaller than the diameter of the heat source, this model represents a good approximation. The average film temperature excess over the bulk oil temperature is given by

$$T = 4Hr_0/3\pi k,$$

where H is the heat generated per unit area per second; r_0 is the radius of the circular heat source; and k is the conductivity of the bearing metal. The thermal gradient over the bearing area is not very large, maximum temperature rise at the center being greater than the average by a factor of 1.18.

An estimate of the rates of shear occurring in the crossed cylinder bearing is of interest because most of the existing evidence for the great viscosity increase of oils when subjected to very high pressures has been obtained in falling body viscometers operating with very low shear rates. An order of magnitude lower limit for rate of shear in the crossed cylinder bearing has been obtained by dividing the linear velocity of the rotor by the clearance limit previously given. At a representative velocity of 300 rpm this lower limit is 200,000 radians per second.

Surface Effects

Since a lower limit on bearing clearance was not established, it is conceivable that oil film thicknesses approached the range where surface induced orientation of the lubricant might result in anomalously high viscosities for an appreciable fraction of the film. Since, however, the lubrication is markedly dependent upon bulk viscosity, the existence of effects arising from differences in surface induced orientation for the several oils are not indicated by the experiments described above.

FIG. 10. Examples of bearing contour maps. (Run 999—run 1484—run 1543.)

Experimental attempts were made to discover effects

TABLE V. Pit depths by interference fringe measurements.

Run	Pit diameter	Plastic pit diameter	Depth by fringe method	Depth of rotor arc
985	0.0333 cm	0.0310 cm	4.57×10^{-4} cm	4.48×10^{-4} cm
992	0.0336	0.0310	4.28	4.54
999	0.0340	0.0310	4.57	4.65
1484	0.0330	0.0310	4.57	4.36
1543	0.0510	0.0426	10.33	10.45

TABLE VI. Insensitivity of friction to operating variables.

A. Effect of viscosity or speed changes[a]

Vis. at 100°F	Speed	Friction coefficient	Bearing pressure
480.4 cs	27 rpm	0.077	151,000 psi
	50	0.084	189,000
	75	0.085	196,000
	800	0.075	215,000
222.5	30	0.068	135,000
	50	0.071	151,000
	100	0.074	180,000
	200	0.077	184,000
69.8	100	0.070	129,000
	200	0.077	138,000
	300	0.084	150,000
	500	0.073	152,000
	800	0.072	147,000
37.2	400	0.066	114,000
	600	0.066	140,000
	700	0.066	142,000
	1100	0.065	142,000

B. Effect of surface finish changes[b]

Vis. at 100°F	Finish	Speed	Friction coefficient	Bearing pressure
69.8 cs	¼μ-diamond	600 rpm	0.062	140,000 psi
		1100	0.059	152,000
	8μ-diamond	500	0.073	152,000
		800	0.072	147,000
222.5	8μ-diamond	30	0.068	135,000
		100	0.074	180,000
		200	0.077	184,000
	3/0 emery	800	0.070	117,000
		1800	0.064	154,000
20.2	8μ-diamond	400	0.085	126,000
		600	0.080	137,000
		800	0.082	140,000
	4/0 emery	800	0.072	131,000
		1800	0.064	114,000

C. Effect of load changes[c]

Load	Friction coefficient	Bearing pressure
23.6 lb	0.067	160,000 psi
31.4	0.072	170,000
47.2	0.061	170,000

[a] Finish 8μ-diamond, 23.6-lb load.
[b] 23.6-lb load.
[c] Finish 3/0 emery, 1800-rpm speed, oil viscosity 222.5 cs at 100°F.

resulting from surface adsorption by introducing into neutral oils acids known to be strongly adsorbed from oil solution. In order not to affect the bulk viscosity measurably, concentrations up to only 0.65 neutralization number (mg KOH per g) were used. This procedure failed to produce any evidence for surface layer lubrication. Pit diameters were measured at operating speeds ranging from the critical speed at which plastic pit size just sufficed for no-wear lubrication with the neutral oil down to lower speeds at which equilibrium pit diameters increased several thousandths of a centimeter. No reduction in pit diameter was effected by the addition of acids as described.

Solidity of High Pressure Oil Films

Although the high pressures in the oil films in the crossed cylinder bearing are maintained by hydrodynamic forces, the films themselves are not necessarily liquid throughout the bearing. The literature of high pressure viscometry contains numerous references to apparent solidification of mineral oils under pressures much lower than bearing pressures reported here.

The probability that the oil films are in fact not liquid throughout the bearing is suggested by friction measurements, which show the coefficient of friction to be independent of speed and nearly independent of viscosity. A compendium of equilibrium friction measurements for a variety of operating conditions is given in Table VI.

Coefficient of friction measurements show considerable variation, differences from run to run of 10 percent being usual and of 20 percent not uncommon. Within such limits, however, the effect of speed variation cannot be detected and the effect of viscosity variation is just discernible. Although the hydrodynamic theory of the crossed cylinder bearing has not been developed, the laws of frictional dependence upon speed and viscosity in journal bearings[12] would hardly lead one to expect such results.

Constancy of friction coefficient with respect to speed is more consistent with a picture of oil film solidity. Whenever load is supported upon a solid lubricant and slippage occurs only within the lubricant, the friction coefficient becomes the ratio between shear strength of the solid and average bearing pressure. In this case the magnitude of friction coefficient has been found to be only slightly dependent upon bearing speed.[13,14]

The slightness of the friction variation with viscosity exhibited in Table VI is not necessarily implied by a solidification picture. For this observed behavior to be consistent with such a picture, it is necessary that the given oils possess substantially the same shear strengths when solidified.

Independent measurements of the shear strength of solidified oils could, therefore, afford a partial check upon the oil film solidification picture. Such measurements have been made under conditions where a hydrodynamic wedge cannot be the mode of bearing pressure development and yield coefficients of friction which agree, within limited precision, with those in the crossed cylinder bearing. These measurements will be described in Part II of this paper.

[12] M. D. Hersey, *Theory of Lubrication* (John Wiley & Sons, Inc., New York, 1936).
[13] P. W. Bridgman, Proc. Am. Acad. Arts Sci. **71**, 387 (1937).
[14] J. Boyd and B. P. Robertson, Trans. Am. Soc. Mech. Engrs. **67**, 51 (1945).

PART II. SHEAR STRENGTH MEASUREMENTS

Shear Strength Apparatus

The experiments on mineral oils described subsequently were performed using an apparatus designed specifically to measure the high pressure shear strengths of solids. In the design of this equipment, the principle of using convex mating surfaces to create known areas of contact and known high average pressures has again been employed.

The basic method used to measure shear strengths is somewhat similar to that used by Bridgman,[13] and later, by Boyd and Robertson.[14] In their apparatus the sample to be tested was first placed between two plane surfaced anvils and then compressed by heavy loads. The torque required to twist one anvil relative to the other was then measured for various normal pressures. In order to avoid the use of a thrust bearing having unknown and presumably variable friction, two such anvil pairs were stacked one upon the other. In this manner, the thrust bearing became a duplicate source of measured torque. The presently described apparatus (shown schematically in Fig. 11 and in completed assembly in Fig. 12) combines this stacked anvil pair technique with the convex mating surface principle in a manner which eliminates the need for large applied loads.

In this apparatus the bearing areas are formed by pressing a hardened steel ball between two plane surfaced anvils until plastic flow of the latter occurs. The spherical depressions thus formed serve the dual purpose of holding the test sample and preserving alignment during subsequent rotation of the ball. The torque required to turn the ball assembly is measured by transmitting the driving force through a resistance strain gauge. As previously stated, with such an arrangement the entire measured torque originates in the bearing areas. Pressures achievable in this apparatus depend on the elastic limit of the anvil material, a range of 30,000 to 250,000 psi being obtained at convenient pit diameter of about a millimeter with common metals and loads up to 350 lb. The aforementioned figures refer to average pressures. There is a pressure gradient over the contact area which will be discussed in the following section.

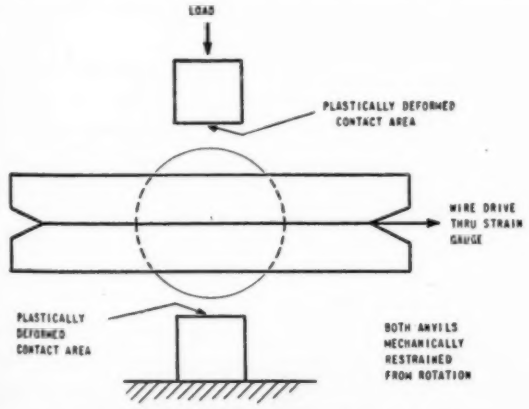

LOAD
PLASTICALLY DEFORMED CONTACT AREA
WIRE DRIVE THRU STRAIN GAUGE
PLASTICALLY DEFORMED CONTACT AREA
BOTH ANVILS MECHANICALLY RESTRAINED FROM ROTATION

FIG. 11. Shear strength apparatus mechanics.

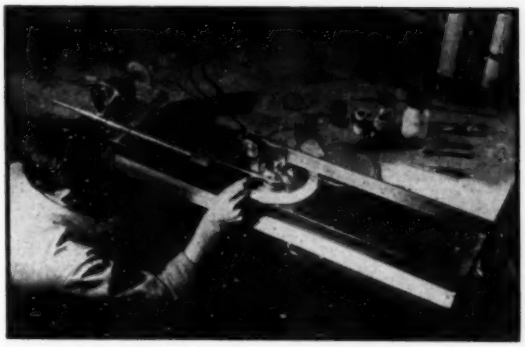


FIG. 12. Shear strength apparatus.

Computation of Shear Strength

The total frictional torque is made up of contributions from shear stresses over the contact area. The integration of these contributions must take into account variations in shear strength which may result from the variation in normal pressure over the contact area.

The pressure distribution in the presence of intervening mineral oil films has been shown to be the same as with dry metal. Interference fringe contour mapping of plastically formed depressions in anvils after load removal showed that no differences in pit contour or diameter result from the interposition of oil. The contours observed were spherical and of larger radius of curvature than the indenting ball.† For this geometry Tabor[10] has demonstrated that the Hertz elasticity equations yield elastic deformations approximating those observed. The pressure distribution can therefore be obtained from the Hertz theory and may be represented by the ordinates of a hemisphere erected on the area of contact.[11]

In order to take into account the effect of variations in shear strength induced by this variable pressure, an assumption must be made regarding the dependence of shear strength upon pressure. Although this dependence is not the same for all substances, its exact nature is not highly critical in affecting integrated average shear strengths. For example, if shear strength is assumed independent of pressure, the resulting integrated value is only 13 percent less than that obtained if shear strength is instead assumed proportional to pressure.

The latter assumption that unit shear strengths vary proportionately with pressure has some experimental justification. For example, both Bridgman[13] and Boyd and Robertson[14] have shown that shear strength is an increasing function of normal pressure for virtually all solids examined. Although strict proportionality between shear strength and pressure is hardly characteristic of each individual solid, such an assumption is probably the most proper generalization from their data that can be made.

† Similar results have been reported by earlier investigators using a different method, F. E. Foss and R. C. Brumfield, Proc. Amer. Soc. Testing Materials **22**, 312 (1922).

TABLE VII. Coefficients of friction of oils at 35,000 psi and 77°F.

Vis. at 100°F	Vis. at 210°F	f*	Rate of rotation
222.5 cs	17.9 cs	0.048	0.23 rpm
		0.048	1.00
		0.050	2.37
220	10.0	0.086	0.23
		0.086	1.00
		0.091	2.37
		0.102	10.0
		0.114	38.9
480	29.7	0.085	1.0
44,000 (extrapolated)	318	0.064	1.0
		0.081	38.9

* Coefficient of friction taken as unit shear stress divided by average normal pressure.

Unit shear strengths have been integrated over the contact areas using this assumption and the simplifying approximation that the spherical contact area is equivalent to its planar projected area. (These areas are equal within a fraction of a percent.) The results, expressed in terms of average coefficients of friction (coefficient of friction taken as the ratio of unit shear strength to normal pressure) are represented by the following equation:

$$f_{\text{average}} = 8T/3\pi RW, \qquad (1)$$

where R is the radius of contact area; W is the applied load; and T is the observed frictional torque for the two contact areas.

Friction Coefficients of Solidified Oils

At moderate pressures in the neighborhood of 35,000 psi, achieved by pairing zinc as the anvil metal with a hardened steel ball, sufficiently viscous lubricating oils behaved in all respects as do accepted solids such as long chain fatty acid soaps, paraffin wax, or the soft inorganic, Ag_2SO_4. Essentially complete separation of the ball and anvil surfaces was effected by the oil even when periods ranging up to 16 hr intervened between load application and ball rotation. After load removal, microscopic observation revealed negligible evidence for the formation of surface junctions or for metal transfer. The microstructure of the zinc anvil surfaces, originally prepared with 3/0 paper, was not altered with the exception that the area within the spherical shell scattered light more strongly than the uncontacted area. This light scattering, however, is also observed when the steel ball is simply pressed into the zinc anvil without lubricant or rotation and, therefore, is not indicative of failure of the lubricant to separate the surfaces.

In order to achieve complete surface separation it was necessary to impart a high degree of polish to the steel ball and to select a sufficiently viscous oil. Deviation toward either rougher surfaces or lighter oils resulted in surface welding, evidenced in irregularly increasing frictional torque and by microscopic observation.

TABLE VIII. Coefficients of friction in crossed cylinder bearing and shear strength apparatus.

Test equipment	Viscosity range	Speed range	Coefficient of friction range
Crossed cylinder bearing	20–480 cs at 100°F	27–1800 rpm	0.059–0.085
Shear strength apparatus	220–44,000	0.23–38.9	0.048–0.114

In addition, attempts were made to ascertain the minimum thicknesses of oil necessary to achieve complete surface separation. Oil layers of calculated average thickness were deposited upon the zinc anvils by evaporating upon a known area a known mass of a light solvent solution containing the oil. By this technique the distribution of oil, after solvent evaporation, is, of course, not uniform, so that precise determinations of minimum thicknesses were not possible. From numerous experiments, using the first oil in Table VII, minimum average thicknesses of 1 to 5 micro-inches (250 to 1250A) were found necessary to secure surface separation of the highly polished steel ball and 3/0 paper finished zinc anvils. The average thicknesses of the oil films sheared in the experiments conducted with this apparatus were therefore in a range at least that great.

As is the case for a solid, frictional torque exhibited by a sufficiently viscous oil is a constant, characteristic of the oil, independent of the angle through which the ball has been rotated and, within broad limits, independent of the rate at which rotation took place. By contrast, if an adhering liquid is contained between two plane parallel disks which are rotated coaxially with respect to each other, hydrodynamic theory predicts that the frictional torque will vary directly with the viscosity of the liquid, directly with the rate of rotation of the disks, and inversely with the separation of the disks. That such is not the behavior of mineral oils at the stated pressures is evident from the results obtained upon the shear strength apparatus. These results are listed in Table VII. These data show that oils differing widely in viscosity at room temperature and atmospheric pressure exhibit comparatively small differences in frictional properties at 35,000 psi. Since the mean separation of the spherical surfaces was not controlled in these experiments and the viscosities themselves are unknown at 35,000 psi, sufficient evidence for solidity cannot be adduced from this frictional behavior alone. However, the insensitivity to rotational velocity also illustrated in Table VII calls for a non-liquid picture in explaining the room temperature frictional behavior of the oils at 35,000 psi.

Friction measurements on mineral oils in two types of apparatus have been reported. In two salient respects these apparatus impose different requirements upon the oils. The crossed cylinder bearing develops high pressures by means of a hydrodynamic wedge and subjects the oil films to high rates of shear; the shear strength apparatus on the other hand traps the oils between two

spherical surfaces at somewhat lower pressures and subjects them to small shearing displacements and low rates of shear. Both apparatus, together encompassing this broad experimental range, are lubricated by intact oil films. In neither apparatus, however, are expected hydrodynamic relationships even qualitatively adequate to explain the frictional behavior observed.

In addition to this mutual deviation from expected liquid behavior, the magnitudes of the friction coefficient measured are sufficiently alike to suggest that both apparatus measure essentially the same property. A composite of coefficients of friction measurements made with both the crossed cylinder bearing and the shear strength apparatus is given in Table VIII. These friction observations taken together with the solidification evidence in the literature of high pressure viscometry have led to a solidified oil film picture as the most plausible to be adduced.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 4 APRIL, 1951

X-Ray Measurement of Short Range Order in Ag-Au*

N. Norman and B. E. Warren
Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received December 26, 1950)

The diffuse intensity from single crystals of AgAu and Ag_3Au was measured with crystal monochromated CuKα radiation and a Geiger counter spectrometer. Measurements were made at a series of points in the (h_1h_20) layer of reciprocal space. The measurements were made at both room temperature and at 160°K to allow a correction for temperature diffuse scattering. For the final annealing temperature of 300°C, the short-range order parameters obtained were $\alpha_1=-0.08$ for AgAu and $\alpha_1=-0.05$ for Ag_3Au.

INTRODUCTION

GOLD and silver are each face-centered cubic with almost identical lattice constants 4.070 and 4.078 kX, respectively. The equilibrium diagram for the Ag-Au system shows complete solid solution over the whole composition range, with no evidence of compounds or superstructures. The question immediately suggests itself as to whether we should expect to find a measurable short-range order in a binary system such as this; where the two atoms are practically identical in size, quite similar chemically, and where there is no evidence for a long-range ordered structure. Guinier[1] has reported that powder patterns of Ag-Au show a diffuse peak characteristic of short-range order, but no attempt was made to get a quantitative evaluation of the short-range order parameters. To obtain numerical values for the short-range order, measurements have been carried out on single crystals with the approximate compositions AgAu and Ag_3Au.

EXPERIMENTAL METHOD

The experimental measurements and the evaluation of the data were carried out following the method used by Cowley[2] on Cu_3Au. In a binary solid solution with atoms A and B present in the proportions m_A and m_B, in addition to the sharp crystalline reflections, there will be a diffuse intensity characteristic of the short-range order given by

$$(I_{eu})_{SRO}=Nm_Am_B(f_A-f_B)^2 \times\sum_j\left(1-\frac{p_j}{m_A}\right)\exp\left[\frac{2\pi i}{\lambda}(\mathbf{s}-\mathbf{s}_0)\cdot\mathbf{r}_j\right]$$

where N is the total number of atoms, $(s-s_0)$ is the diffraction vector, and p_j is the probability of finding an A atom at a position r_j from a B atom. In the face-centered cubic lattice, the position of one lattice point relative to another can be represented by

$$\mathbf{r}_j=la_1'+ma_2'+na_3'$$

where lmn are certain integers and $a_1'a_2'a_3'$ are each half of the usual cubic axes $a_1a_2a_3$. In terms of the $a_1'a_2'a_3'$, the corresponding reciprocal vectors are $b_1'b_2'b_3'$, and the diffraction vector $(s-s_0)$ can be represented by

$$\left(\frac{s-s_0}{\lambda}\right)=h_1b_1'+h_2b_2'+h_3b_3'$$

where $h_1h_2h_3$ are continuous variables. The short-range order parameter α_{lmn} is defined by

$$\alpha_{lmn}=1-(p_{lmn}/m_A), \qquad (1)$$

where α_{lmn} is the order parameter for neighbors at a distance lmn, and p_{lmn} is the probability of finding an A atom as an (lmn) neighbor of a B atom.

Introducing the above quantities, the expression for the short-range order diffuse intensity becomes

$$(I_{eu})_{SRO}=N_{m_Am_A}(f_A-f_B)^2I_{SRO}', \qquad (2)$$

* Research sponsored by the ONR under contract N5-ori-07832.
[1] A. J. Guinier, Proc. Phys. Soc. (London) **57**, 310 (1945).
[2] J. M. Cowley, J. Appl. Phys. **21**, 24 (1950).

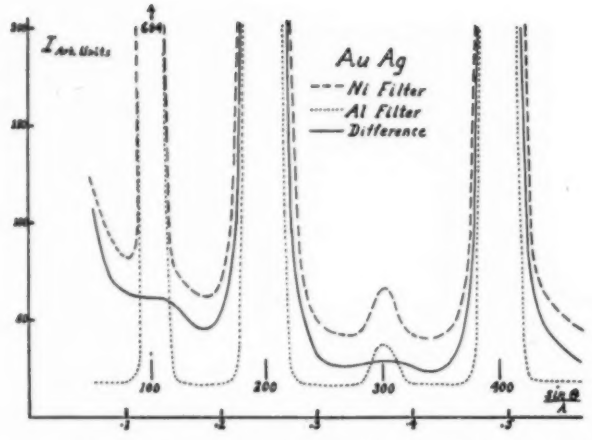


FIG. 1. Measurements along the $h_1 00$ line for AgAu using a quartz $(10\bar{1}1)$ monochromator with CuKα. Correction for $\lambda/2$ by a balanced Ni-Al filter.

where

$$I_{SRO}' = \sum_l \sum_m \sum_n \alpha_{lmn} \times \exp[2\pi i(lh_1 + mh_2 + nh_3)]. \quad (3)$$

As seen from (3) the corrected short-range order intensity I_{SRO}' is periodic in reciprocal space, with periods $b_1'b_2'b_3'$ which are double the reciprocal vector lengths $b_1b_2b_3$ corresponding to the usual cubic axes. The short range order parameters α_{lmn} are the coefficients of the fourier series for I_{SRO}' and can be evaluated in the usual way by

$$\alpha_{lmn} = \int_0^1 \int_0^1 \int I_{SRO}'(h_1h_2h_3) \times \exp[-2\pi i(lh_1 + mh_2 + nh_3)]dh_1dh_2dh_3. \quad (4)$$

Single crystals of the alloys AgAu and Ag_3Au were prepared by heating to above the melting point in a graphite crucible and cooling at about 13°C per hour. A steady stream of nitrogen prevented oxidation, and a temperature gradient was maintained by cooling the lower end of the crucible. Single crystals about 1 or 2 cm in size were obtained. These were oriented by Laue patterns and cut parallel to the (100) plane. After polishing and etching the crystals, they were annealed for 1 week at 450°C, 1 week at 400°C, 2 weeks at 350°C, and finally, 4 weeks at 300°C.

The experimental procedure was similar to that used by Cowley. The primary beam fell on a (100) face of the crystal and the diffracted radiation was measured with a Geiger counter. Three degrees of freedom in the motion of the crystal allowed the reciprocal vector to sweep through any desired part of reciprocal space, at the same time maintaining equal angles between the primary and diffracted beams and the face of the crystal. A bent quartz monochromator focused CuKα radiation on the face of the crystal, and the divergence of the primary and diffracted beams was set at about 3 degrees to give sufficient intensity for the measurements.

With the extended face technique, the observed *SRO* intensity can be written

$$I_{SRO} = \frac{KN_0P}{\mu}\frac{(I_{eu})}{N}SRO,$$

where $(I_{eu})_{SRO}$ is given by Eq. (2), N_0 is the number of atoms per unit volume, μ is the linear absorption coefficient, P is the polarization factor, and K is a constant. The constant K can be determined from a suitable standard material; we have found that the high angle scattering from paraffin is an excellent standard for putting diffuse intensities upon an absolute basis.

The total observed intensity is made up of 7 terms: the sharp crystalline reflections due to λ and $\lambda/2$, the short-range order diffuse intensity, the Compton modified scattering, the temperature diffuse scattering, fluorescence radiation, and air scattering. The crystalline reflections due to λ give the intense sharp peaks at the fundamental points in reciprocal space; in these regions the diffuse *SRO* intensity has its minima. A quartz monochromator using the $(10\bar{1}1)$ planes diffracts an appreciable component of $\lambda/2$ from the continuous spectrum. Since there is one of these $\lambda/2$ reflections by a fundamental point at each of the maxima of the *SRO* intensity, and the extraneous peak is much stronger than the diffuse intensity we are measuring, a balanced Ni-Al filter (balanced at $\lambda/2$) was used for each reading. Figure 1 shows how bad the $\lambda/2$ component may be when measuring weak diffuse peaks such as those at 100 and 300. The measurements were taken along the $h_1 00$ line in reciprocal space. The weak peak at 100 is too

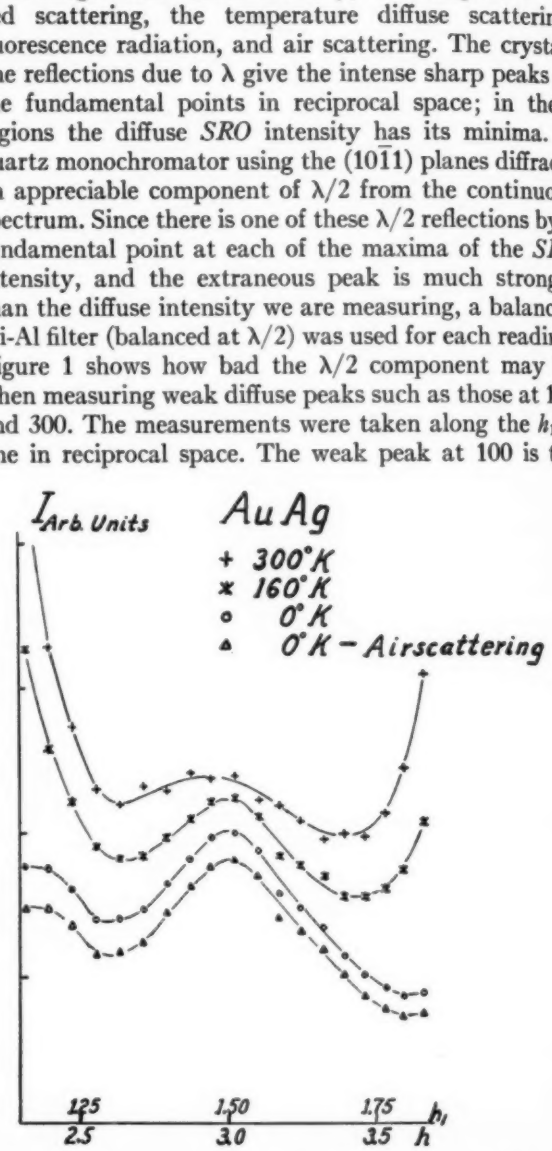


FIG. 2. Measurements on AgAu in the region of the 300 superstructure point illustrating the corrections for temperature diffuse and air scattering.

dependent upon the correction to be reliable, but at 300 the correction was considered safe.

The Compton modified scattering can be computed theoretically and subtracted as soon as the measured intensities are standardized in absolute units. To eliminate the quite considerable amount of temperature diffuse scattering which has its largest values close to a fundamental peak, readings were taken at room temperature and at 160°K, and assuming a linear dependency for the temperature scattering, the effect could be extrapolated out. The low temperature was obtained by blowing prepurified nitrogen through a thermos bottle of liquid nitrogen directly onto the surface of the crystal, the steady stream of dry nitrogen preventing the moisture in the room from reaching the surface and forming ice. Fluorescence radiation was cared for by the filters, and the air scattering was measured by removing the crystal.

The curves of Fig. 2 prove the existence of short range order in AgAu, and show the magnitude of the temperature diffuse and the air scattering. The readings are taken along the h_100 line and cover the region of the short range order peak at 300. Each point is the difference of the Ni and Al filter readings. The upper curve was taken at room temperature and the next at 160°K. Assuming linear dependence on T, and extrapolating to 0°K gives the third curve. Subtracting the measured air scattering gives the lower curve. The small rise at the left is probably due to a small error in correcting for large intensities at 200. The final curve shows the pronounced peak at the 300 position which is to be expected if short range order is present. It is interesting to notice that the temperature diffuse scattering largely masks the peak at room temperature, and that it does not really stand out until the temperature scattering has been removed.

Figures 3 and 4 show the measured intensity distribution in the h_1h_20 plane of reciprocal space for AgAu and Ag_3Au. In this paper position in reciprocal space is represented by the variables $h_1h_2h_3$, but the crystalline reflections are labelled by the usual indices hkl. Thus the fundamental reflection (400) occurs at $h_1=2$. For convenience the observed intensities in arbitrary units have been divided by $P(f_{Au}-f_{Ag})^2$. Considering the symmetry about the [100] and [110] axes, the intensity distribution for the whole plane can be found up to $\theta=60°$. The peculiar forms of the contours, and the ridges such as the ridge running across from (400) to (220) are due to temperature diffuse scattering. In fact, the existence of short range order peaks is not clearly seen on Figs. 3 and 4 until after the temperature diffuse scattering is subtracted.

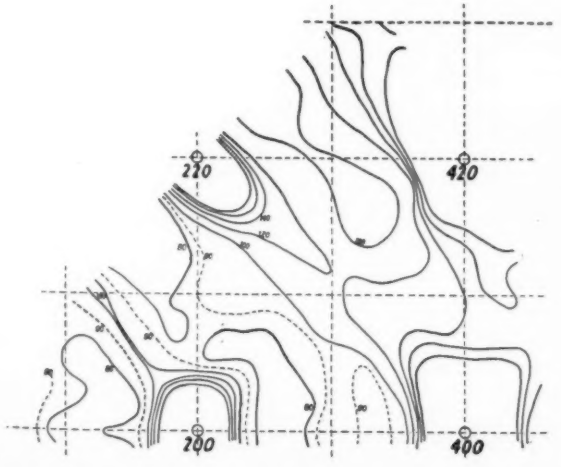


FIG. 3. Contour plot showing the measured diffuse intensity for AgAu at 27°C in the h_1h_20 plane of reciprocal space.

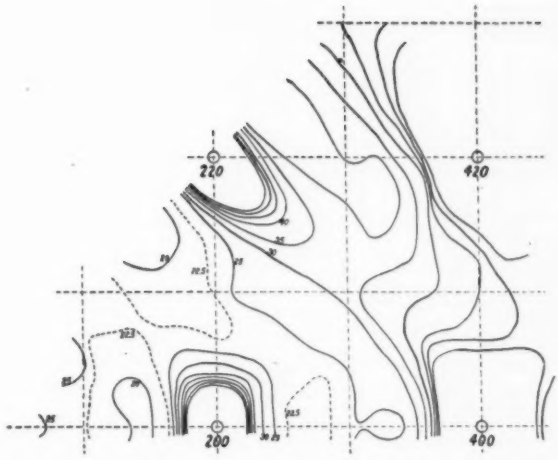


FIG. 4. Contour plot showing the measured diffuse intensity for Ag_3Au at 27°C in the h_1h_20 plane of reciprocal space.

RESULTS

For measurements in the h_1h_20 plane only, h_3 is zero and Eq. (3) becomes

$$I_{SRO}'(h_1h_20)=\sum_l\sum_m\left(\sum_n \alpha_{lmn}\right)\times\exp[2\pi i(lh_1+mh_2)]. \quad (5)$$

The fourier coefficients of the intensity distribution thus give a quantity $A(lm)$ where

$$A(lm)=\sum_n \alpha_{lmn}=\int_0^1\int I_{SRO}'(h_1h_20)\times\exp[-2\pi i(lh_1+mh_2)]dh_1dh_2. \quad (6)$$

The integral of Eq. (6) is evaluated for a few combinations of the integers l and m, giving quantities $A(lm)$ which represent certain combinations of the *SRO* parameters α_i where i designates the shells of neighbors ($i=1$ nearest neighbors, etc.). Since the short range order in Ag-Au is small, the α_i decrease rapidly and only the first few need be considered. For this case measurements throughout a volume of reciprocal space are not necessary. Considering terms only to the 10th nearest

TABLE I. Experimental values of $A(lm)$.

	AgAu	Ag_3Au
$A(00)$	+1.050	+1.170
$A(10)$	−0.165	−0.105
$A(11)$	−0.086	−0.045
$A(20)$	+0.013	+0.011
$A(21)$	+0.003	−0.002

neighbor, the expressions for $A(lm)$ are

$$A(00)=\alpha_0+2\alpha_2+2\alpha_8, \quad A(21)=2\alpha_3+2\alpha_7,$$
$$A(10)=2\alpha_1+2\alpha_5, \quad A(22)=\alpha_4+2\alpha_6,$$
$$A(11)=\alpha_1+2\alpha_3+2\alpha_9, \quad A(30)=2\alpha_5+2\alpha_9,$$
$$A(20)=\alpha_2+2\alpha_4+2\alpha_{10}, \quad A(31)=\alpha_5+2\alpha_7.$$

The values of $A(lm)$ were computed from the intensity distribution in the square limited by 200, 300, 310, and 210. Intensities were measured at points obtained by dividing this square into 10×10 intervals. The values of $A(lm)$ were then obtained by approximating the integral by a summation and using the Lipson and Beevers strips for the calculation. The values obtained are listed in Table I.

Assuming the α_i small enough to be neglected beyond $i=2$, and making no use of the $A(00)$ values since they contain $\alpha_0=1.0$ and are very sensitive to the accuracy of standardizing the curves, the values for α_1 and α_2 listed in Table II were obtained.

TABLE II. Values of short-range order parameters.

AgAu	Ag_3Au
$\alpha_1=-0.08$	$\alpha_1=-0.05$
$\alpha_2=+0.01$	$\alpha_2=+0.01$

TABLE III. Comparison of values of α_1 and number of A atoms about a B atom.

		Complete disorder	Perfect order	Measured
AgAu	α_1	0.00	−0.33	−0.08
	No. of A neighbors	6	8	6.48
Ag_3Au	α_1	0.00	−0.33	−0.05
	No. of A neighbors	9	12	9.45

DISCUSSION

For the parameter values listed in Table II, the negative sign for α_1 and the positive sign for α_2 indicate a tendency for unlike nearest neighbors and like second nearest neighbors. Each atom has 12 nearest neighbors and for complete disorder the average numbers will be 6 Au, 6 Ag for AgAu, and 3 Au, 9 Ag for Ag_3Au. For maximum short-range order, corresponding to perfect long range order, the numbers of nearest neighbors of a Au atom are 4 Au 8 Ag for AgAu and 12 Ag for Ag_3Au. From the definition of α by equation (1), it follows that α_1 is zero for complete disorder and $-\frac{1}{3}$ for maximum order for both the AB and the A_3B compositions. The significance of the α_1 values listed in Table II is given by the comparisons of Table III.

Both samples were annealed down to a temperature of 300°C so that the measured order corresponds to this temperature or perhaps a somewhat lower temperature. For AgAu the measured order is 24 percent of the way from disorder to perfect order and for Ag_3Au it is 15 percent. It is interesting to note that the measured values of α_1 for AgAu and Ag_3Au have a ratio 0.08:0.05, which is roughly that predicted by the Cowley[3] theory. The computed values of α_1 for CuAu and Cu_3Au at 770°C which are shown in Cowley's Fig. 5 are in the ratio 0.14:0.11. An approximate value of T_c for Ag_3Au can be obtained by extending the α_1 curve of Cowley's Fig. 3 to higher values of T/T_c. For Ag_3Au $\alpha_1/\alpha_1^\circ = -0.05/-0.33=0.15$ and this corresponds to $T/T_c=3.6$. Assuming that the measured order corresponds to 300°C, the critical temperature for Ag_3Au becomes $T_c=160$°K. This is so low that a long-range ordered form of Ag_3Au would not be expected. The nearest neighbor interaction energy of Cowley's theory can be obtained from T_c. The value obtained is $V_1=89k$ as compared with $V_1=358k$ for Cu_3Au. On the basis of thermodynamic measurements of the activity of Ag in Au-Ag solid solution, Professor Averbach has calculated the free energy of mixing to be −257 cal per mole corresponding to $V_1=129k$. The existence of short-range order in the Ag-Au system, in spite of the close similarity in size of the two atoms, is in keeping with the negative deviation from Vegard's Law shown by a plot[4] of lattice constant against composition.

One of us (N. Norman) was assisted in this work by fellowships from the Royal Norwegian Council for Scientific and Industrial Research and from the Norwegian Ministry of Ecclesiastical Affairs and Education.

[3] J. M. Cowley, Phys. Rev. **77**, 669 (1950).
[4] C. S. Barrett, *Structure of Metals* (McGraw-Hill Book Company, Inc., New York, 1943), page 203.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 4 APRIL, 1951

Kerr Cell Photography of High Speed Phenomena*

E. M. Pugh, R.v. Heine-Geldern, S. Foner, and E. C. Mutschler
Department of Physics, Carnegie Institute of Technology, Pittsburgh, Pennsylvania
(Received November 15, 1950)

Visible light photographs have been obtained of metal jets squirted from the lined conical cavities of high explosive charges. Since these jets travel through air at nearly meteoric velocities (7 to 12×10^5 cm/sec), their front ends are heated to incandescense and vaporized. The remainder of the jet is relatively nonluminous and is photographed by synchronizing a Kerr cell shutter (exposure time <1 μsec) and an exploding wire light source (peak intensity 4×10^8 candlepower) with the phenomenon. Schlieren type photography cannot be used because of the luminosity accompanying the phenomenon. Detail is obtained in the photographs by using a large lens with a seven-in. focal length. To maintain the effective aperture at $f/4$ requires a large Kerr cell which in turn requires a 25,000 volt pulse for its operation.

The technique is used in studying the action of these jets on steel, water, glass, and Plexiglas. The technique is also used in studying the propagation of shock waves in solids and liquids when these phenomena are due to the action of the metal jets or to the more direct action of explosives.

INTRODUCTION

THE metal jets which are squirted from high explosive charges with lined cavities[1] travel at velocities exceeding 7000 m/sec. Since they are relatively nonluminous compared to the intense luminosity of the accompanying explosion, their photography by visible light presents some unusual problems. For that reason, previous attempts at photographing these jets have been unsuccessful, the only satisfactory results having been obtained by flash x-ray methods.[2]

Visible light silhouette photographs have now been obtained by synchronizing a very brilliant light source (electrically exploded wire of peak intensity 3 to 5×10^8 candlepower) and a Kerr cell shutter with the phenomenon. To obtain detail of the fast moving jet at an object-to-lens distance of 3 to 6 feet, the exposure time must be one microsecond or less. Since fast (and therefore large grain) film must be used, the negative must be large enough to preserve detail. This is accomplished by using an aerial lens having a focal length of seven inches and an aperture of $f/2.5$. This aperture is reduced slightly by the Kerr cell, so that pictures are taken at an effective aperture of $f/4$.

Figure 1 shows a schematic and simplified diagram of the experimental setup. The explosion takes place in a bomb proof separated from the observation room by a heavy, reinforced concrete wall. A port hole protected by bullet proof glass allows the camera to take pictures of phenomena occurring in the bomb proof. The phenomenon is silhouetted against a diffusing screen of tracing paper illuminated from the back by the light of the electrically exploded wire. The Kerr cell with its crossed polaroids transmits light only during the microsecond interval when a 25-kv pulse is applied to the electrodes of the Kerr cell.

JETS TRAVELING IN AIR

Early attempts to photograph the high speed jets from shaped charges in visible light were usually

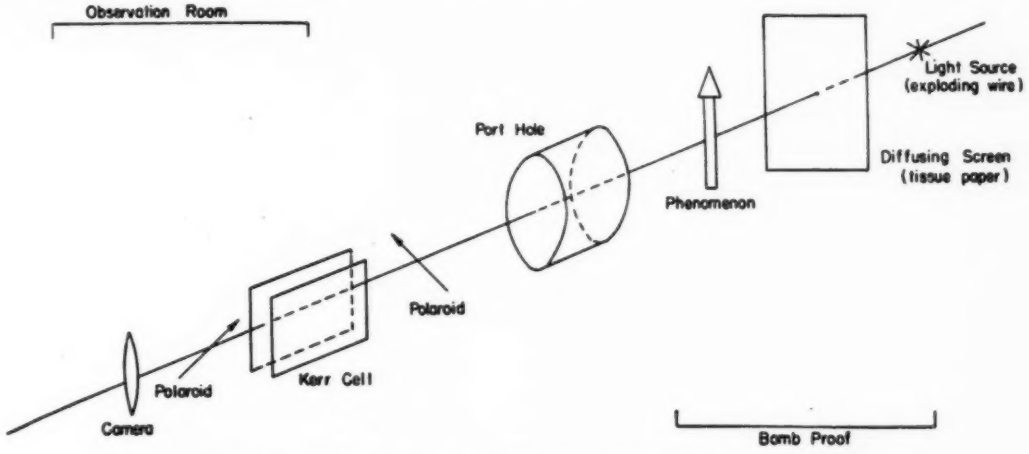


FIG. 1. Simplified schematic diagram of experimental setup used for Kerr cell photography.

* This work was carried out under a research contract with the Office of the Chief of Ordnance, Contract No. W-36-061-ORD-2910, and has been released for publication by the Office of Public Information, Department of Defense.
[1] Birkhoff, MacDougall, Pugh, and Taylor, J. Appl. Phys. **19**, 563 (1948).
[2] J. C. Clark, J. Appl. Phys. **20**, 363 (1949).

FIG. 2. Jet from a copper cone with flat apex. Only a small portion of the jet is visible, the rest being surrounded by an opaque shroud of spall particles which are vaporized by their rapid motion through air.

unsuccessful,[3] showing even less jet than can be seen in Fig. 2. While a small portion of the jet can be seen near the upper portion of that picture, most of the jet is obscured by an opaque shroud. In fact the pictures did not show any jet at all until the light source was made so brilliant that some light was transmitted by the nearly opaque cloud surrounding the jet and until the exposure time was made short enough to stop the motion.

When the tip of the shroud is magnified by several diameters, it appears to consist of a large number of high speed particles traveling nearly parallel to the jet and vaporizing continuously by their rapid passage through air. This conclusion was verified by taking a photograph of a jet passing through an evacuated glass tube. At pressures below 10^{-3} mm Hg, no shroud at all is seen, the particles being too small to be visible in the photograph, and no metal vapor being formed in the absence of air.

[3] In 1943, investigators at the Explosive Research Laboratory in Bruceton, Pennsylvania (MacDougall, Paul, Eyster, and Messerly) did obtain some photographs of jets in a vacuum, using a flash bulb made of high explosive. These pictures have never been shown in an unclassified publication.

The jet shown in Fig. 2 resulted from a copper cone with a flat apex. (In the commercial production of these cones it is impractical to make liners truly conical down to a pointed apex.) However, the lack of a pointed apex was found to be responsible for the production of most of the shroud seen in Fig. 2, for when nearly perfect brass cones were made on a lathe, the photograph shown in Fig. 3 resulted. There almost all of the jet is visible, only a small portion at the tip of the jet being obscured by the metallic vapor (incandescent near the tip) produced by its rapid motion through air. (F. Zwicky and F. Whipple have proposed using these jets for the study of "artificial meteors." They have not previously been photographed at close range.)

Figures 4 and 5 are intended to give some idea of the physical setup used for the photographs shown so far. They show the setup and the action picture, respectively, of two metal jet charges detonated simultaneously by equal lengths of primacord fuse running to a single "distributor" charge.

In a later paragraph pictures of jets penetrating certain target materials are discussed. To facilitate interpretation of those pictures, two other topics are considered first: the detonation of explosive charges, and the shock waves and cracking damage produced in

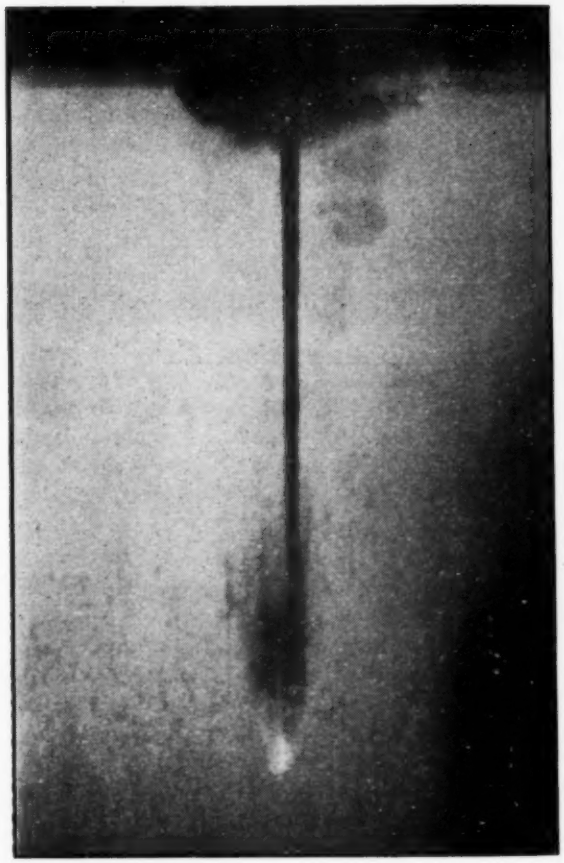


FIG. 3. Jet from a machined brass cone with truly conical apex. The shroud accompanying the jet shown in Fig. 2 has been nearly eliminated.

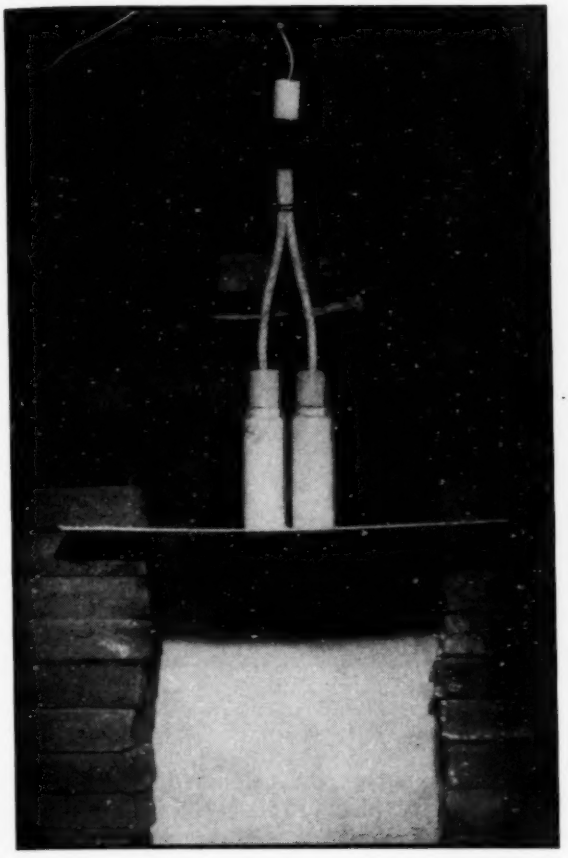


FIG. 4. Two metal jet charges standing side by side on a sheet of cardboard and fused with primacord for simultaneous detonation.

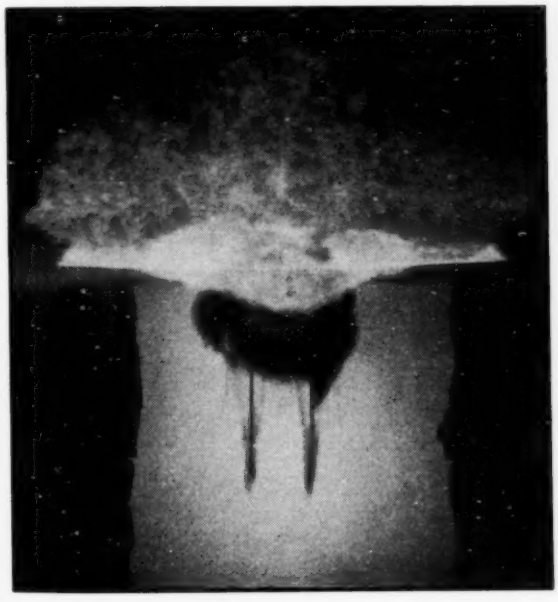


FIG. 5. Jets resulting from the simultaneous detonation of the two charges shown in Fig. 4.

FIG. 6. Detonation of a cylinder of pentolite. The detonation wave is traveling downward at a velocity of 7510 m/sec.

various materials when placed in contact with a detonating charge.

DETONATION PHENOMENA

The following photographs were obtained with both pentolite, a high explosive having a detonation velocity

FIG. 7. Simultaneous detonation of two pentolite cylinders. Note region of Mach interaction between the two expanding cones of explosive products.

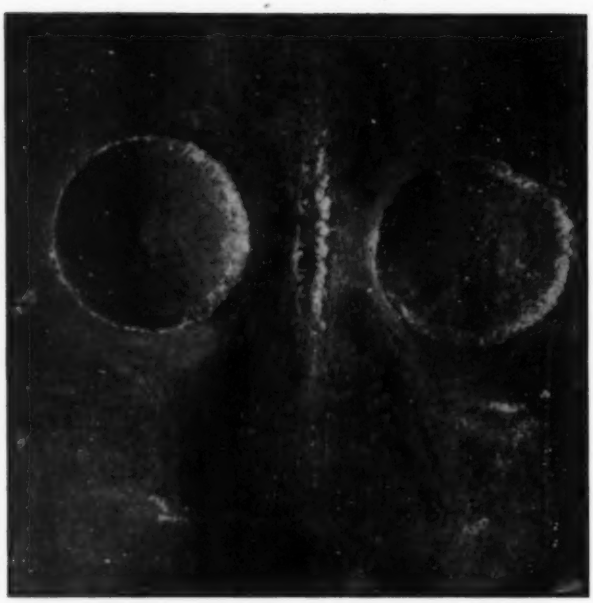


FIG. 8. Appearance of the mild steel plate on which the two charges shown in Fig. 7 had been standing.

of 7510 m/sec, and with primacord (a fast explosive fuse), having a detonation velocity of 6420 m/sec.

Figure 6 shows a one-inch diameter cylinder of pentolite in the process of exploding, the detonation wave proceeding downward. The upper portion of the cloud of explosive products is caused by the initiating mechanism of the charge. Measuring the angle between the conical envelope of explosive products and the charge axis (37.5°), and using the known detonation velocity of 7510 m/sec for pentolite, the calculated initial velocity of radial expansion of the explosive products is approximately 5800 m/sec.

Interference effects between two envelopes of explosive products intersecting at an angle are shown in Figs. 7 and 9. In Fig. 7 two one-inch diameter pentolite cylinders were placed on a steel plate with axes parallel,

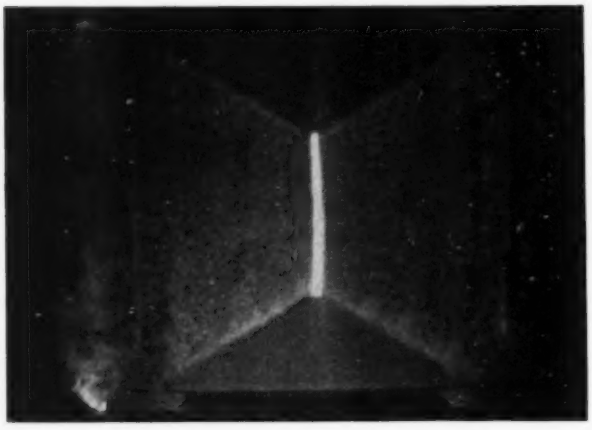


FIG. 9. Cylindrical pentolite charge detonated simultaneously from both ends.

FIG. 10. Detonation of primacord under water. Detonation wave proceeded downward at a velocity of 6420 m/sec.

$2\frac{1}{2}$ in. between centers, and detonated simultaneously. A region of very high luminosity, caused by Mach

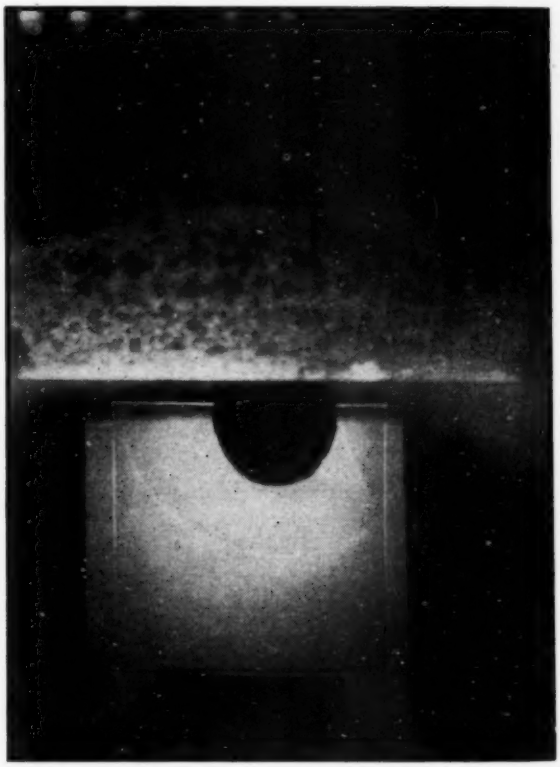


FIG. 11. Detonation of a small pentolite charge standing on the top face of a $6\times6\times1\frac{1}{4}$-in. slab of Plexiglas. Note original shock wave, reflected tension wave and dark region of shattering.

interference, is clearly visible. The photograph of the 4×4-in. mild steel plate on which the charges had been standing is shown in Fig. 8. Besides the two circular indentations produced by the flat ends of the charges, a well-defined cut was produced in the plate by the Mach stem which indicates the order of magnitude of the pressure within it. The depth of indentation (3 to 4 mm) due to this interference effect was approximately equal to that produced by the charges themselves. For Fig. 9, a long cylinder of pentolite was detonated simultaneously from both ends. Here a ring-shaped luminous region of Mach interference was produced along the circle of intersection of the two expanding cones of explosive products.

SHOCK WAVES AND CRACKING IN TARGET MATERIALS

The Kerr cell method of high speed photography has been applied to studies of shock waves and fracturing in various materials. The shock produced by a length of primacord detonated under water is shown in Fig. 10, the dark central region consisting of the (relatively) slowly expanding explosive products. Due to the air-water interface at the top of the tank, a definite curvature of the shock wave, corresponding to that predicted by a Huygens construction, is visible. From such photographs it is found that this shock wave propagates in water at a velocity ranging from 2200 m/sec close to the primacord to about 1700 m/sec some distance away from it. At still longer distances, the shock wave velocity would probably approach the velocity of sound in water (1450 m/sec). A similar picture (not shown) of a pentolite charge detonated under water indicates an initial shock wave velocity of 3200 m/sec.

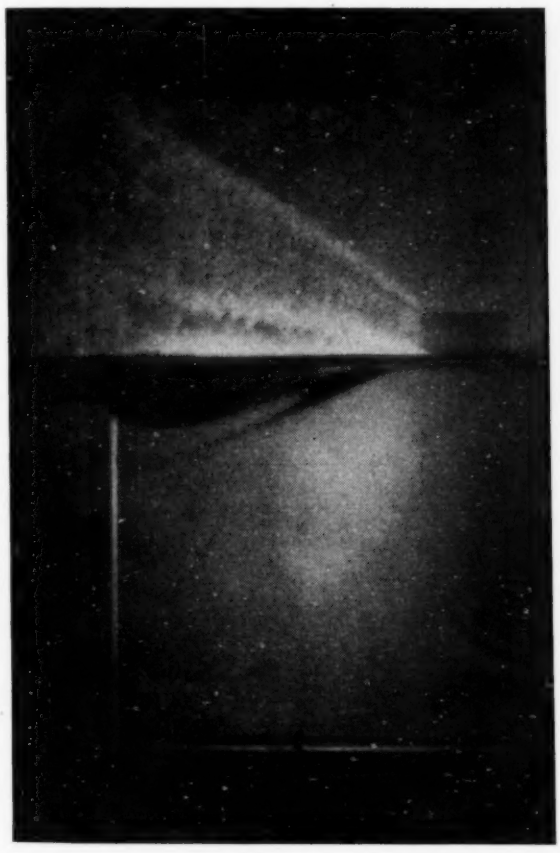


FIG. 12. Detonation of a pentolite cylinder laid horizontally along the upper face of a piece of Plexiglas.

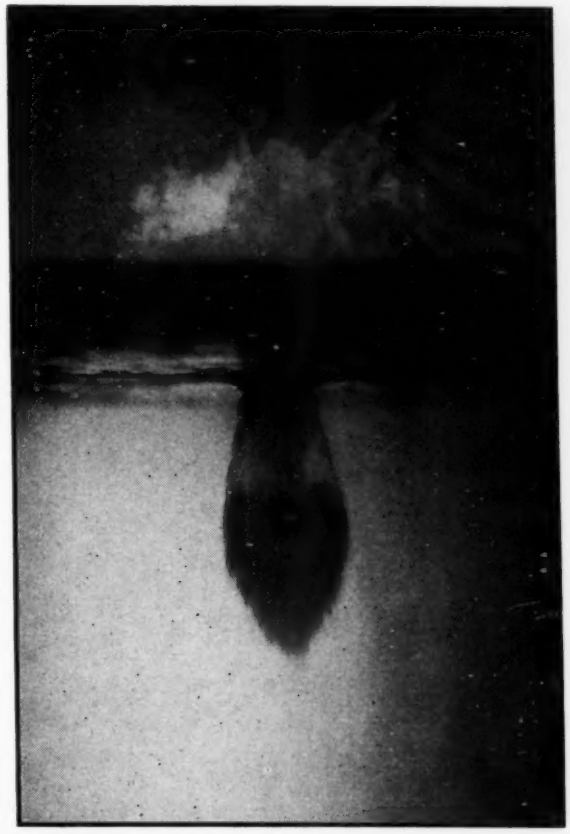


FIG. 13. Metal jet perforating a ½-in. steel plate. Note intense spalling from the bottom surface of the plate.

The shadow technique described so far is useful for observing high speed phenomena in transparent solids. For this purpose, Plexiglas targets of dimensions 6×6 ×1¼ in. were used.

A cylinder of pentolite was placed on such a Plexiglas target and detonated to produce the results shown in Fig. 11. The cylinder of explosive was placed to the right of the center on the top 6×1¼-in. face, so that one observes a reflected tension wave from the right-hand face in addition to the original compression wave. One also observes the extent of shattering at the time the exposure was made.

Since the detonation velocities of most explosives are well known, shock wave velocities in transparent solids can be determined conveniently by laying an explosive charge on one face of the material and measuring the angle between the supporting face and the shock wave. Figure 12 shows the shock wave produced in this

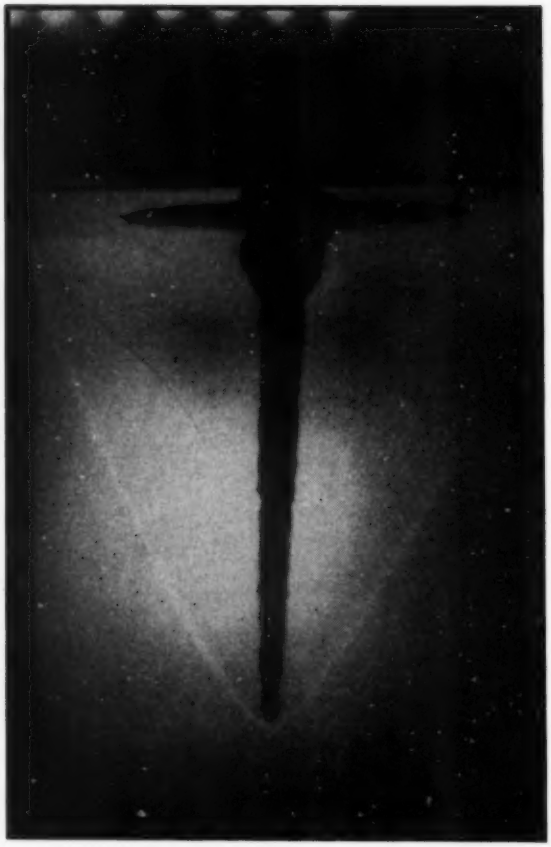


FIG. 14. Metal jet penetrating water. The jet itself is not visible, because it is surrounded by a region of intense turbulence.

manner by pentolite laid on the upper face of the Plexiglas target. The shock wave varies from 3400 m/sec for the most intense shock, to 2400 m/sec as the shock intensity decreases. This lower velocity agrees well with the sound velocity measured by Protzman.[4]

The last five figures show photographs of metal jets in the process of penetrating or perforating various target materials. In the case of Fig. 13, the target material was a steel plate $\frac{1}{2}$-in. thick. At the top of the picture the cloud of explosive products can be seen moving downward with the rear part of the jet. Just above the steel plate, near the point where it is struck by the jet, another cloud consisting of a particle spray coming from the target can be seen. The pear-shaped cloud underneath the plate, consisting mostly of particles spalled from the steel plate, is sufficiently dense to obscure the tip of the jet, while the remainder of the jet is now passing through the hole in the plate without interference.

Figure 14 shows a jet shot into a glass tank filled with water. The velocity of penetration of water under these circumstances is known to be about 5000 m/sec. Since the shock wave makes an angle of nearly 30° with the jet, its velocity must be close to 2500 m/sec, or about 1.7 times the velocity of sound in water. The jet itself is not visible in the picture, since it is completely surrounded by a region of violent turbulence in the water.

A jet penetrating a piece of Plexiglas of dimensions $6\times6\times1\frac{1}{4}$ in. is shown in Fig. 15. The area above the tip of the jet is opaque because the shock wave has reached the front and back surfaces and the resulting reflected tension wave has shattered the Plexiglas in that region. The picture indicates that no shattering is produced by the intense compression wave at the tip of the jet.

Figure 16 shows a jet penetrating into a $1\frac{1}{4}$-in. thick piece of glass after it has perforated a similar slab of plexiglas, the two slabs being separated by a $\frac{1}{2}$-in. air space. The Plexiglas plate shows a reflected tension wave traveling upward from the bottom surface, in addition to a rather complicated cracking pattern. The jet inside the glass is surrounded by a transparent region which in turn is bounded by a dark line. This dark line cannot be a shock wave since the shape of a shock wave would be conical rather than cylindrical. It seems probable that the transparent region consists of molten glass which has been melted by the high pressure produced by the tip of the jet: energy considerations indicate that the glass in that region should be heated to a temperature between 1000 and 1500°C. If this is correct, then the black line surrounding the jet repre-

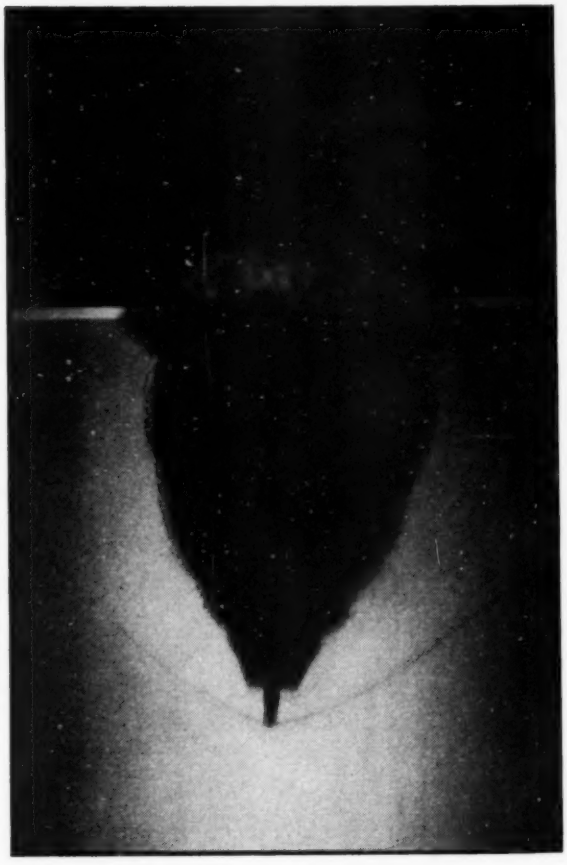


FIG. 15. Metal jet penetrating through Plexiglas.

[4] T. F. Protzman, J. Appl. Phys. **20**, 627 (1949).

FIG. 16. Metal jet penetrating into a glass plate after having perforated a similar slab of Plexiglas, separated from the glass by a ½-in. air space.

FIG. 17. Jet penetrating a glass plate at an angle of incidence of 20° to the surface of the plate.

sents a boundary between glass in its solid and its liquid phase.

Figure 17 shows the jet entering a glass target at an angle of 20° to the target surface. The boundary between the liquid and solid phases is again evident. Cracking has begun at the rear surface directly opposite the point where the jet entered the glass, indicating that the shock wave has reached that surface and has been reflected. This photograph clearly demonstrates that the jet is not deflected when it impinges at a glancing angle on the surface of a hard material like glass. Experiments also show that it is not deflected by the hardest steel.

Positive Coaxial Cylindrical Corona Discharges in Pure N_2, O_2, and Mixtures Thereof*

CHARLES G. MILLER† AND LEONARD B. LOEB‡
University of California, Berkeley, California

(Received November 2, 1950)

Dc corona studies were made on coaxial cylinders with inner cylinder at positive potentials in pure N_2, pure O_2, and mixtures of 1 percent, 10 percent O_2 in N_2 and on clean air from 27-mm to 760-mm pressure. Observations of current-potential relations from field intensified ionization currents to as near breakdown as possible were made noting thresholds for current transitions, pulsed discharges, and other oscillographic as well as visual appearances. Positive corona in pure N_2 does not exist as such. Predischarge Townsend currents by photon action on the cathode condition the outer Ni cathode until it spontaneously develops a glow discharge focusing in a spot on the anode wire. The discharge sets in at high potential and persists with lowering by hundreds of volts. Increase in current gives multiple spots and irregular current fluctuations. At lower pressure potential overshoot gradually disappears and the discharge begins as a regular glow discharge. Positive coronas with 1 percent O_2 show burst pulses at threshold, these being longer in duration (1000 μsec) than in air because of spread down the wire. At lower pressure burst pulses lasted still longer. Breakdown streamers were observed at lower pressures where the applied potential was adequate. The Townsend equation for current in coaxial geometry applies above threshold both for N_2 and N_2+1 percent O_2. In air the positive corona sequence showing only burst pulses at threshold and breakdown streamers were observed. Burst pulses lasted two to four times those with 1-mm diameter point to plane corona in air owing to spread along the wire, which was less extensive than in 1 percent O_2. In pure O_2 the high absorption of photons ionizing the gas led to pre-onset streamers and no burst pulses at thresholds above 200-mm pressure. Below 200 mm streamer pulses and burst pulses were observed and below 50 mm only burst pulses were observed. The streamers above 200-mm spread down the whole length of the tube by transparency of O_2 to photons liberating electrons from the Ni cathode cylinder.

1. INTRODUCTION

THE existing literature on the action of coaxial cylindrical coronas in their application to Geiger counters and on a larger scale in engineering studies of coronas is extensive. No really comprehensive study of the *nature of the mechanisms* active has ever been undertaken over sufficient pressure ranges, using pure free electron and electron attaching gases under such conditions as to enable these phenomena to be correlated with the extensive studies of the group at the University of California on the basic mechanisms as derived from point-to-plane coronas. In fact, no really comprehensive study, even of the point-to-plane corona, over an extended pressure range in air has been completed to date. In consequence, the large proportion of those working on Geiger counters do not know they are dealing with a burst pulse corona, nor that in pure N_2, for example, there is not properly speaking a corona but a concentrated form of glow discharge that could never make a suitable counter unless fouled with oxygen. It is the purpose here to describe and interpret briefly the very wide range of phenomena which the much needed study under the title of this paper has revealed. A parallel study in air on a standardized point-to-plane system is now nearing completion in the hands of H. W. Bandel in this laboratory.

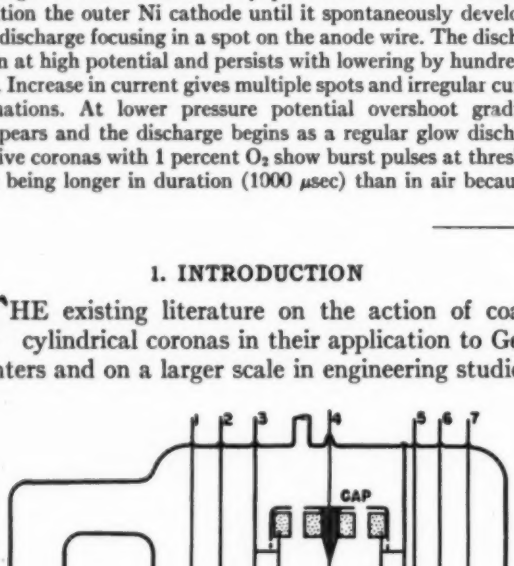


FIG. 1. Diagram of apparatus.

2. APPARATUS AND METHODS

The discharge tube (see Fig. 1) consists of a nickel collecting cylinder, 28.5-mm inside diameter and 5.55 cm in length, placed accurately coaxial with nickel guard cylinders of the same diameter. The spacers and supports are fashioned of quartz or of Pyrex, in different tubes, with no apparent difference in insulation quality. The central collecting electrode has two longitudinal viewing slots, 1 mm×45 mm milled at 90° separation, for the purpose of viewing phenomena simultaneously at right angles through a pair of telemicroscopes. It has, in addition, a hole 1 cm in diameter, screened with 60-mesh nickel screen, through which ultraviolet light or other

* This work was sponsored by the Research Corporation of New York and in part by the ONR.

† Currently at the University of California at Santa Barbara.

‡ The experimental work was entirely done by Dr. Miller. Owing to the complex nature of many of the phenomena much of the interpretation fell to the lot of Professor Loeb.

radiation, may be admitted to the sensitive volume through a quartz window in the glass envelope (not shown in Fig. 1).

By means of two tungsten filaments inside the glass envelope and placed opposite the collector and guard cylinders the nickel cylinders can be brought up to red heat in a short time for outgassing by electron bombardment. These filaments also serve to scavenge the last traces of oxygen for the part of the program dealing with nitrogen free oxygen.

The center platinum filament, diameter 0.006 in., is kept accurately coaxial with the cylinder by means of guides. This center platinum filament can be outgassed by passing a current through it when under high vacuum. It is so mounted that a spring under tension keeps it always taut through repeated outgassing and bakeout.

The glass jacket has a quartz window, joined by a graded seal to the jacket, opposite the screened opening of the collector electrode. One other feature of the glass jacket which proved useful is a glass chimney for circulating the gas by convection. When air or oxygen-nitrogen mixture was being studied, the chimney saved much time over that needed for dispersal of discharge products by diffusion.

The chamber was mounted in a furnace so arranged that it could be baked out at 425°C for from six to twelve hours, while the metal parts were kept glowing by electron bombardment, until the pressure indicated on the McLeod gauge showed less than 10^{-5} mm pressure while hot. Maintenance of this pressure at the temperature of 425°C was taken to mean that the bulk of the adsorbed gas had been removed both from the discharge tube and the remainder of the glass system. All the glass not included in the furnace was vigorously flamed during the bakeout period.

The gas purifying system consisted of a Cu and a CuO train for the purpose of removing oxygen and hydrogen respectively from the tank gases admitted. Pure oxygen was prepared according to the method of N. E. Bradbury.[1] The gases, above atmospheric pressure in the tube leading from the pressure reduction valve of the tank to the system, were admitted to the discharge tube through a glass capillary filler that restricted the flow to approximately 100 cm^3 per min, and then through the hot Cu or CuO train followed by three liquid nitrogen traps. In the case of oxygen at partial pressures exceeding 150 mm of mercury the cold traps were immersed in dry ice-acetone baths. The entire system was flushed repeatedly with the gas to be used before the final filling.

The high voltage source which was to prove most useful consisted of two 5000-v, continuously variable, reversible polarity, well-regulated power supplies. The corona currents were measured on a Dolazalek electrometer, using a set of internally consistent but not absolutely calibrated S. S. White high resistors, ranging from 2×10^{12} to 10^8 ohms, the electrometer having its highest stable sensitivity of 3150 mm/v at 1.25 meters. Higher currents were measured either by shunting the electrometer with suitable resistors, or by using a microammeter. The readings from a separate microammeter in the guard circuit proved helpful in interpreting the results, as will be noted. A cathode-ray oscilloscope pick-up resistor, variable from 100 to 500,000 ohms was placed in the collector-ground return to pick up the signal. The insulation for the lowest current leads was of amber or of freshly cast sulfur.

The applied high voltage was determined by measuring with a type *K*2 potentiometer the potential drop across a 1000-ohm wire wound resistor placed in series with a 12-megohm wire wound resistor, in order to limit the current drawn from the supplies.

The platinum filament was protected by 5 and 10 ma Littel fuses, rated to open in 0.1 msec.

3. DEFINITIONS

A *discharge* current exists when owing to a potential difference an ionization current independent of triggering electrons flows across any gas separated electrode system. When the system is at low potential differences, the negative and positive ions created by natural ionization are collected. As the potential is raised primary ionization by collision first increases the current proportional to the natural ionization as a *field intensified* current. Beyond this point further increase may make secondary ionizing mechanisms active which lead to self-sustaining discharges. Some of these by their visible manifestations were called coronas. These studies have revealed the frequent presence of self-sustaining currents of low order with ill-defined thresholds under present techniques. They usually show no luminosity and precede the appearance of the more notorious coronas. These discharges appear to be ones in which the secondary mechanism consists of a photoelectric liberation from the cathode. They will hereafter be called *Townsend discharges*. With increase in potential, or occasionally with time, these currents will "condition" the cathode leading to amore efficient secondary mechanism with much increased current flow. Again, the increase in potential about a positive point may initiate secondary mechanisms *in the gas* leading to much enhanced currents. These relatively sharp increases in current by one or many orders of magnitude, over limited potential ranges, with or without luminosity, will be termed *corona thresholds* with various characteristic designations. Owing to statistical fluctuation and especially to space charge inhibitions at the thresholds, the discharges appear intermittent or interrupted making exact determination of the threshold potential difficult. With still further increases in potential, or better, current the marginal space charges are removed more and more rapidly so that at some more or less clearly defined potential the discharge goes from intermittent to continuous. This transition will be designated as the *onset of*

[1] N. E. Bradbury, Phys. Rev. **40**, 508 (1932)

steady[2] *corona*. By the time onset currents are reached, most coronas give visual manifestations. At positive corona threshold current pulses of low amplitude and durations of tens of microseconds sometimes appear and are called *burst pulses*.[3]§ They are caused by *lateral* spread of the discharge over the anode surface by photoelectric ionization in the gas and are *the characteristic phenomena* in Geiger counters. They cover the anode with a thin, tightly adhering glow. At or above *positive corona threshold*, and *below onset* of steady corona, very sharp, heavy pulses of less than 10^{-7} sec duration appear. These are called *preonset streamers*.[4] They project the discharge from the anode into the gas some tenths of a mm to mm by intensely localized photoelectric ionization of the gas. They are luminous and appear as brushes of various forms beyond the burst pulse glow on the point. They often reappear just before the gap breaks down to a spark. They are then longer and more intense interrupted discharges. These are called *breakdown streamers*[4,5] and are individually clearly visible.

At negative corona threshold the discharge may be a continuous discharge, either concentrated or diffuse. Irrespective of dispersion, these negative discharges will be called glow discharges. Other negative discharges at threshold appear as intermittent or pulsed discharges. Where such pulses have a duration of the order of microseconds, the pulse will be called the *Trichel pulse*[6] after its discoverer. These pulses are very heavy glow discharges striving towards an arc over. They fail to arc as they are choked off by intense negative ion space charge formation in the low field regions. All cathodes surfaces being under positive ion bombardment, tend to be unstable and their behavior depends on past history.[7] With adequate current densities, or time of exposure to positive ion bombardment, cathodes "*clean up*" or "*condition*." They then become more efficient secondary emittors leading to overvolted gaps. Such conditions cause threshold potential values to be above the sustaining potentials for an operating discharge. This condition will be called the *overshooting of a threshold*. Where such actions as "clean up" lead to very heavy currents, which but for limiting resistors or other agencies would go to arcs, the thresholds will be described as *breakdown thresholds*. These imply a rather complete electrical breakdown of the gas. Negative currents of high density will always concentrate in space at narrow spots on cathodes as Loeb has shown.[8] The spot will be the narrower the higher the current density. This property of cathode spots will be noted at many points. At about 10^{-2} amp/cm^2 the diameter is of the order of 0.02 cm diameter. The diameter is relatively insensitive to decrease of the pressure down to 100 mm beyond which it increases. Such spots repel other cathode spots. Further reference to this property of cathode spots will not be made in interests of brevity. It must be recalled whenever spots are mentioned.

There are a number of quantitative relations setting the conditions for the various thresholds. These are gathered at this point in the interest of brevity and will be referred to by number.

Glow Discharge Threshold[9]

$$\gamma \exp \int_a^r \alpha dx = 1. \tag{1}$$

The expression $\exp \int_a^r \alpha dx$[10] is the Townsend integral for the *avalanche* of cumulative electron ionization in the nonuniform field ranging from the region of low ionizing efficiency a to the cathode radius. The symbol γ represents the chance that a positive ion striking the cathode will liberate an electron. This normally has values of the order of 10^{-2} to 10^{-3} for a "conditioned" surface.

Townsend Discharge Threshold[10]

$$\frac{\eta\theta g}{\alpha} \exp \int_a^r (\alpha - \mu) dx = 1. \tag{2}$$

Here μ is the absorption coefficient of the gas for photons, η is the number of photoelectrically active photons that can act on the *cathode* created per electron of the avalanche, g is the fraction of the photons which reach the cathode in the face of geometrical limitations and θ is the efficiency of these photons in liberating electrons from the cathode.

Burst Pulse Threshold[11]

$$\beta f \exp \int_a^r \alpha dx = 1. \tag{3}$$

Here f is the fraction of photons created by the avalanche electrons which are capable of liberating a photoelectron *in the gas*, β is a geometrical factor which determines the fraction of the photo-electrons which are

[2] G. W. Trichel, Phys. Rev. **55**, 382 (1939) also A. F. Kip, Phys. Rev. **54**, 139 (1938).
[3] A. F. Kip, Phys. Rev. **55**, 549 (1939).
§ E. J. Lauer has recently shown that with sensitive high speed oscillograph and small collecting capacity using positive wire in pure H_2 individual avalanches triggered by photoelectric action at the cathode cylinder may be observed. With larger collecting capacity a succession of such avalanches that eventually chokes off by space charge or by statistical fluctuation yields an integrated pulse shape that simulates burst pulses but should be termed *Townsend pulses* since they emanate from photoelectric action at the cathode instead of in the gas.
[4] L. B. Loeb and A. F. Kip, J. Appl. Phys. **10**, 142 (1939).
[5] L. B. Loeb, *Fundamental Processes of Electrical Discharge in Gases* (John Wiley and Sons, Inc., New York, 1939), p. 533. Also, L. B. Loeb and Leigh, Phys. Rev. **51**, 199 (1937).
[6] G. W. Trichel, Phys. Rev. **54**, 1078 (1938).
[7] L. B. Loeb, Phys. Rev. **76**, 255 (1949).
[8] L. B. Loeb, J. Appl. Phys. **19**, 882 (1948).
[9] See reference 5, p. 410.
[10] See reference 5, p. 380.
[11] L. B. Loeb, Phys. Rev. **73**, 798 (1948).

liberated in a portion of the gas where they can help propagate the discharge.

Streamer Threshold[12]

This is given by a most complicated relation, for the derivation of which see Loeb and Wijsman, which reads,

$$\tfrac{1}{2}f_1f_2[\alpha^{-1}\mu\rho^3/(\delta)^{\frac{1}{2}}] \times\int_\rho^\delta r^{-1}\exp(-\mu r)\left[\exp\int_\rho^r \alpha^1 dr^1\right]dr=1. \quad (4)$$

The onset of *steady burst pulse corona* has not quantitatively been defined before. It may be set as:

$$\beta^1\nu f^1\exp\int_a^r \alpha dx=1. \quad (5)$$

Here f^1 is the fraction of the electrons in the last avalanche of a dying burst pulse, choked off by space charge, that liberate a photon capable of removing an electron from the Ni cylinder. β^1 is the geometrical and other chance that they reach the cathode cylinder without absorption and ν is the chance that the electron is liberated by the surface.

4. PURE N_2-POSITIVE WIRE

The current-potential curves observed in pure N_2 with positive wire are shown for a range of pressures from 617 to 27 mm in Fig. 2. The currents rise relatively steeply with potential from the field intensified ionization current of about 10^{-12} amp, to a critical value of between 10^{-8} and 10^{-9} amp. For the higher pressures the rise is more gradual. The rise terminates in an abrupt jump in current to a very high value of about 10^{-5} or 10^{-4} amp with little change in potential. Below 99 mm the transition is more irregular. Once the higher pressure coronas break into the high current value, reducing potential does *not retrace the curve* but leads to a rather irregular gradual decline of the current well *below the potential* which caused the break. These curves can be followed down with decreasing potential until the corona goes out and the current drops abruptly to values less than 10^{-8} amp. This *overshoot* of potential, or "hysteresis" effect, is always associated with a conditioning action on the cathode, in this case the outer Ni cylinder wall. The cathode-ray oscillograph showed no pulses or indication of any but a small steady current, which is not recorded in the method of observation used, as long as the jump did not occur. As soon as the jump‖ from 10^{-9} amp occurred, the oscilloscope showed very irregular current oscillations of considerable amplitude about a high

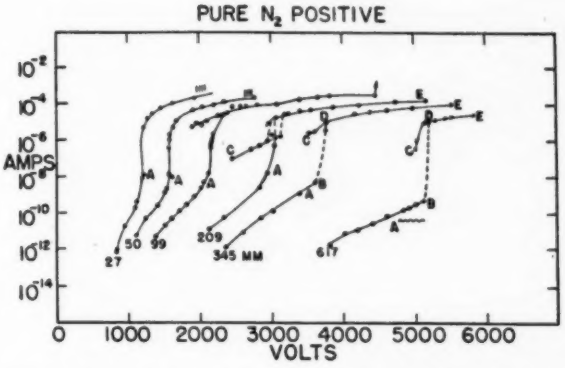


FIG. 2. Current potential diagrams pure N_2 positive point. Note steep rise of current in region marked A. Note abrupt transitions at point marked B representing appearance of corona with multiple spots at D. Note the decrease in current with decreasing voltage in region dc representing overshoot. Points marked E indicate limit of current given by source potential indicated. Arrow on curve for 99-mm pressure indicates spark. Vertical lines above curves at 27 and 50 mm indicate breakdown streamers preceding spark.

constant background over time intervals of the order of 0.1 to 1 msec as in Fig. 3A. These are not the well-known burst or streamer pulses. The nature of the corona discharge under these conditions was at once revealed by the visual manifestations. Below the current jump no luminosity was seen. Coincident with the current jump to 10 μa there appeared *multiple bright spots* on the anode wire shown in Fig. 3B. These spots were in continual motion. Increasing the potential multiplied the spots, decreased the fluctuations and smoothly increased the average current about which the fluctuations appeared. As the current decreased below 10 μa the current became very unsteady and the spots went out one by one, eventually leaving one single luminous spot with a current of 2.5 μa, or thereabout. Further reductions in potential led to the extinction of the spot and to a drop in current by a factor of 5 or more. For example, at 617 mm the current was 10 μa at 5050 v with multiple spots; the single spot appeared at a current of 8.5 μa and persisted with a current of 0.5 μa at 5040 v to the central section of the cylinder and 0.1 μa to the guard rings. Below this the spot went out and the current went to 10^{-10} amp with the oscilloscope showing a slight ripple at that current. If the potential was raised immediately, the spot sometimes appeared at a slightly lower potential than before, but usually required the same overshoot of 5140 v at 617 mm or more. Just before extinction with one spot and 0.5 μa current, the oscilloscope pattern was quite steady indicating a concentrated glow discharge. The ratio of length of guard cylinders to collector was 0.85. Above 10 μa with multiple spots on the wire, the current to the guard cylinders was about half of the current to the collector. The appearance at offset was a fine spot on one side of the wire as drawn in Fig. 4A, the spot remaining stationary. This spot resembled those observed by Weissler with very pure N_2 using a point.

The same sequence appeared at lower pressures down

[12] L. B. Loeb and R. J. Wijsman, J. Appl. Phys. **19**, 797 (1948).

‖ While it is the potential that governs this behavior, the transitions often occur within such small changes in potential that its value is not too significant, except in order of magnitude. On the contrary, the current changes are so abrupt and marked that it will be the custom to use this coordinate in discussion and description.

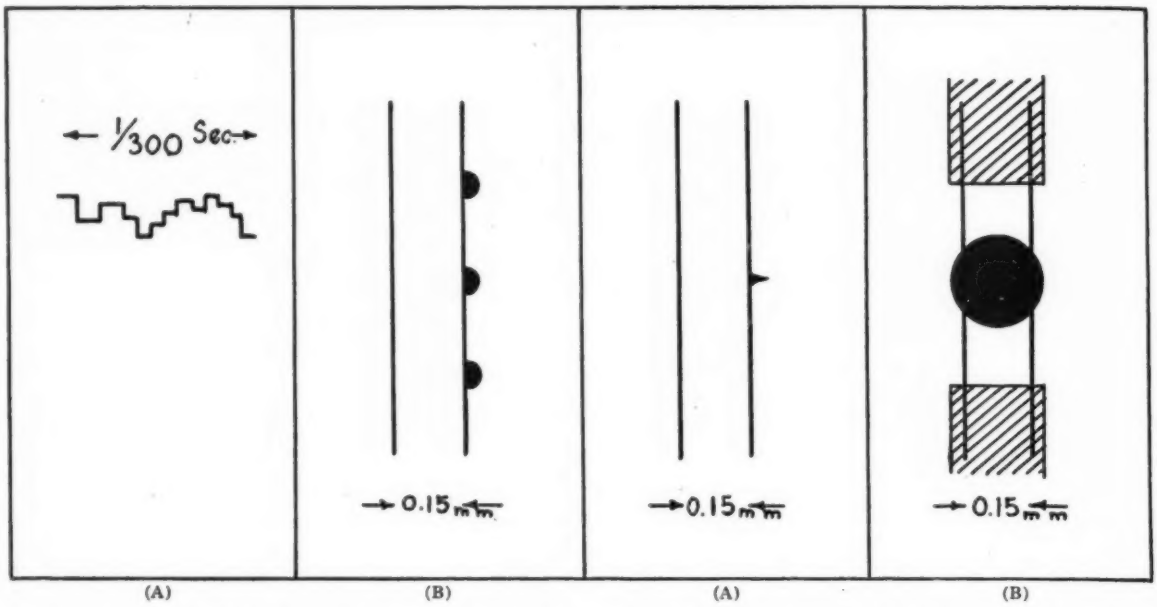


FIG. 3. Figure 3A represents the irregular pulses accompanying fluctuations with multiple spots on wires shown in Fig. 3B.

FIG. 4. Figure 4A represents single spot with decreasing current just before offset. Figure 4B represents single spot and velvety glow on wire at 99 mm before offset.

through 99 mm; only the overshoot was progressively less at reduced pressure and the current jump became less. While there was overshoot at 99 mm, there is a suspicion that if the potential had been raised very gradually, there would have been no overshoot, which was the case at 50 and 27 mm. At 99 mm and 7 μa the whole wire began to manifest a uniform glow instead of the spot. At 15 μa a single spot appeared superposed on the glow, although there was a dark space between glow along the wire and the spot glow as seen in Fig. 4B. At 50 mm all overshoot disappeared and the curves were reproducible in both directions with a uniform glow and no spots to 100 μa. Both the 27-mm curves, which were analogous to the 50 mm, and the 50-mm curves show characteristic breakdown streamers before breakdown at 10^{-4} amp. Results analogous to those noted for 27 mm were observed at 12 mm but the data were not recorded.

It is clear that this system of very pure N_2 and clean Ni metal cathode surrounding the positive wire leads to a sequence of phenomena which *in no way are related to the regular positive wire coronas and to Geiger counter action.*¶ Notable particularly are the rather steep rises in current with potential from ionization currents of 10^{-12} amp up to about 10^{-8} amp, which were not usually observed in this study. The general distributions of current between guard rings and central section below the sharp current jump, with the appearance of spots and an overshoot in potential at the jump, all indicate that in this system there is a smooth transition at relatively low voltage from a field intensified ionization current to a self-sustaining Townsend discharge at potentials well below that of the large current jump. Both pure N_2 and H_2, as Weissler[13] has shown in point to plane corona discharge, are very transparent to photons. The photons are created near the positive wire by the cumulatively ionizing converging electrons as they move into the high field region. The photons are not absorbed in the gas and even down to energies as low as some 4.5 ev, the work function of clean Ni, they can cause photoelectric emission from the cylinder walls. The cylindrical geometry conserves these photons. The photons persist until practically all of them lead to electron emission from the Ni cylinder, which has a field at its surface of about 200 v/cm. Thus, following the condition set by Eq. (2) at near 10^{-11} amp, a Townsend discharge of low current density is created and this furnishes the rather large increases of currents noted before the jump. Not enough is known about the values of $\eta\theta g/\alpha$ needed to initiate the discharge. Experimentally, no changes marking the threshold of the Townsend discharge are noted. The slight statistical fluctuations in going from field intensified current to Townsend discharge are not revealed by the crude current measuring devices. If the value of the potential at the beginning of Townsend discharge be chosen as that of offset or spot extinction, which is certainly high, the value $\eta\theta g/\alpha$ is of the order of 10^{-3}. The Townsend discharge probably set in at a lower potential. It spread down the cylindrical system owing to the diffusion of photons. Increase in potential increased the current and, also, the potential gradients near the cathode. If the filament is not accurately coaxial, it can happen that

¶ The pre-threshold currents observed indicate the presence of a Townsend discharge of the type recently identified by Lauer and termed Townsend pulses. These were not resolved by the large capacity and slow sweep oscilloscopes used in this study in 1947.

[13] G. L. Weissler, Phys. Rev. **63**, 96 (1943).

being nearer one side of the outer cylinder than the other, the discharge will concentrate along one side of the wire feeding to a linear element on the negative cylinder surface opposite. In addition to the photons that are pretty generally distributed along the cylinder, the positive ions created in the neighborhood of the wire begin to strike in restricted localities of the negative cylinder wall in ever increasing numbers and with ever increasing energy. These gradually remove adsorbed gas film from the Ni cylinder. Obviously, such clean up will be irregular and different at different points of the cylinder. Thus in course of time, the time being less, the higher the current, the Ni cylinder will clean up in spots. When this clean up has progressed far enough at some point, the secondary emission from the surface of the cylinder under positive ion bombardment locally gives more secondary electron current by positive ion bombardment than do the more diffuse photons. That is, the threshold given by Eq. (1) is activated at spots with a value of γ very much larger than $\eta\theta g/\alpha$. These spots are then overvolted. The current emitted from the local spots on the Ni cylinder opposite points on the filament of intense ionization will lead to the sudden increase in current by orders of magnitude.[7] At this point the wire breaks into a multiple spot glow. The irregular current bursts at different spots, interference between spots, the wandering character of the discharge on the Ni cylinder wall associated with variable initial conditioning, give the fluctuating oscilloscope pattern observed. Such patterns are not unusual with overloaded multiple spot discharges.[6] Ultimately further increases in current overload the initial spots and more increase in current can only come through increase in spot number. With these come less statistical fluctuation, steadier and increased currents. Reducing potential cuts out the less efficient spots and gradually reduces the discharge to just one spot.

Apparently the spots are deactivated on standing very shortly after the current ceases. They must then be cleaned up anew by increasing the background Townsend discharge. Attempts were made to calculate a value of γ from the relation $\gamma \exp \int_a^r \alpha dx = 1$ for the overshoot and for the offset. The overshoot potentials gave values of γ of the order 6×10^{-4} which is certainly too low. The offset potentials gave better values of γ but these varied rapidly with pressure as they should owing to back diffusion. It must be concluded that the potential for spot appearance or disappearance is determined by the *current density needed to start a clean up of the Ni surface and the time.* The offset potential is that required to give a local current density sufficient to *maintain* a clean surface despite the diffusion of whatever agency deactivates the spot as well as supports the glow discharge.

At lower pressures the current required before the threshold is reached is progressively higher. This could follow from the increase in diffusion and spread of the discharge areas with pressure reduction. Thus, larger currents are needed to give the local current densities of positive ions needed for clean up. Under these conditions the density of ionization on the wire produces enough luminosity to be seen. At still lower pressures, the current densities never reach high enough values to clean up a spot. With higher X/p ratios and the rapid ion as well as photon diffusion, the current can adequately be carried by the wire and cylinder as a whole with a uniform glow. Then the Townsend discharge probably involving both Eqs. (1) and (2) only is observed.

The light at the anode wire surface comes from the anode fall of potential and is the well known anode glow characteristic of glow discharges and arcs which in essence this discharge resembles. Thus, the anode spots are quite luminous and follow the vagaries in localization and wandering of the cathode spots causing them. As current increases with increasing field, the luminosity extends further into a gap from the wire as would be expected.

If the steady diffuse glow discharge at low pressures without spots is governed largely by the positive ion space charge limitation of current, as postulated earlier by Townsend[14] and later by Werner and others, the application of the rough equation of Townsend,

$$V(V-V_0)=\frac{piR^2}{2k_i}\log R/r \qquad (6)$$

should yield proper values of the positive ion mobilities in pure N_2. Here V_0 is the starting potential of the corona, which is essentially the point of large current jump, and V is the actual potential applied, i is the current per cm length of wire, r and R are the cylinder radii, p is the pressure, and k_i is the ionic mobility. For data at 50-, 27-, and 12-mm pressure, the mobility reduced to 0°C and 760 mm was 3.1, 2.9, and 3.0 cm^2/v sec. The conventionally accepted value of the positive ion mobility in pure N_2 lies between 2 and 3 cm^2/v sec. This is satisfactory agreement with such a rough relation.

With very heavy currents over the whole tube, near breakdown at 27- and 50-mm pressure, enough N_2 dissociates to N to permit photoionization in the gas and the appearance of positive breakdown streamers.

This very interesting behavior, however, disappears at once if the least trace of O_2 or other contaminant leading to photoionization in the gas appears. Under these conditions preonset streamers may occur, and burst pulses at once appear and *yield a Geiger counter.* The circumstance cited illustrates how very careful one must be about prejudging phenomena when conditions are slightly altered, i.e., extreme purity in this case makes an obvious Geiger counter-like system fail to show any counting action whatever.

5. POSITIVE WIRE SLIGHTLY IMPURE N_2 AND N_2+1 PERCENT O_2

As might be anticipated from what preceded, even the slightest impurity can alter the picture observed with

[14] J. S. Townsend, Phil. Mag. **14**, 83 (1914). Also, S. Werner, Z. Physik **90**, 356 (1934); **92**, 705 (1934).

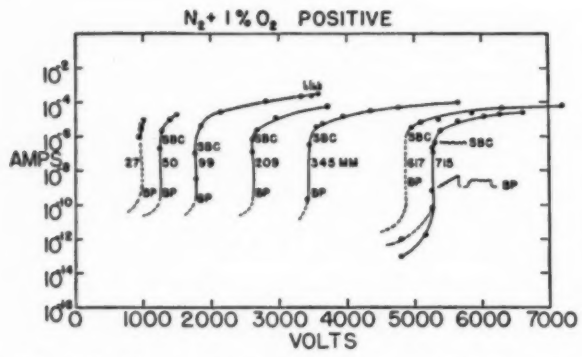


FIG. 5. Positive corona N_2 with one percent O_2. Note abrupt jump of current with burst pulse at BP and the onset of steady pulse corona at SBC. Note appearance of breakdown streamers at end of 99-mm curve.

pure N_2. By accident, flashing one of the auxiliary tungsten filaments which had not previously been outgassed, produced enough contamination to alter observations. The *negative* corona thresholds, current values and behavior were not seriously altered. The positive corona started from a fairly high background current, but there was no overshoot. With the abrupt increase in current burst pulse corona characteristic of the Geiger counter regime was observed. As current increased above the threshold, the tightly adhering velvety glow characteristic of the lateral spread of the burst pulse corona propagating down the wire by photoelectric processes in the mixed absorbing gas was observed. It persisted up to 50 μa at 209 mm and to 20 μa at 345 mm. Above these currents and the corresponding potentials the spots characteristic of gas clean up and a secondary electron liberation from the Ni cylinder by positive ion bombardment occurred superposed on the burst pulse glow. Thus, the contamination was not adequate to prevent the spot breakdown at higher gas pressures, voltages and current densities.

The addition of 1 percent of O_2 to the N_2 produced the set of curves shown in Fig. 5. There the striking changes in mechanism are revealed by the curves. In the region below threshold the currents were less, with indications of only a small Townsend photoelectric contribution above the field intensified currents. Unfortunately, the curves were not followed down to low enough potentials to reveal much detail. The absence of much Townsend discharge attributable to photoelectric effect at the cylinder wall as in pure N_2 is shown by the dashed lower part of the curve at 715 mm. This was obtained by increasing the amount of initial ionization using a radioactive source of increased strength outside. However, the transparency of the gas is such that there was some Townsend discharge as noted by comparison with the curves in the air. These curves were taken with no current limiting resistor so that the *resistance controlling the discharge* was primarily caused by the *ionic space charges in the tube*. Instead of currents rising to 10^{-8} amp as with pure N_2, the currents rose to somewhere around 10^{-10} to 10^{-9} amp and abruptly jumped to between 0.1 and 1.0 μa. Simultaneously with this, the characteristic burst pulse pattern was observed on the oscilloscope, with intermittent bursts lasting to 10^{-3} sec. The duration of these bursts before extinction in the region above threshold, but below onset, is of the order of 10 to 20 times those observed with point to plane gap of similar design.[3] This is to be expected since on a wire the superficial corona which builds up its own extinguishing positive ion space charge can spread laterally into virgin territory along the wire. In contrast, at a point it is confined to a limited area. Furthermore, with more diffuse photoionization the choking positive space charge builds up more slowly. Decreasing pressures down to 50 mm increases the duration of the burst pulses in the preonset region from 1 to 6 msec. This is caused by the increased diffusion and need for a greater number of ions to build up the space charge density to a point where it can choke off the pulse.

The intermittent burst pulse or Geiger counter regime ranges in current from 10^{-9} to 10^{-7} amp but the potential spread of the counting plateau is exceedingly small. In this region the current magnitude, i.e., frequency of bursts, depends on external triggering.[2] The burst pulse threshold is indicated by symbols BP on each curve. The *onset* of steady burst pulse corona is indicated by the symbols SBC on each curve.[2] It begins at around 0.1 to 1 μa. The condition for steady burst pulse corona onset is set by Eq. (5). This follows since the duration and frequency of the bursts, as these increase with potential, must finally reach a point where the next burst is triggered by photo-electrons from the Ni cylinder walls released by photons from imprisoned or other radiation of the preceding burst pulse. The photoelectric propagation of the discharge along the wire at low pressures may account for the spread, as has been assumed in many Geiger counter studies. In N_2 with 1 percent O_2 the Ni cylinder with low work function is also very effective in the spreading of the discharge in the transparent gas. This is in agreement with other speculations as to Geiger counter action.[7] It must be emphasized that the mechanisms active vary radically with conditions such as transparency of the gas. Thus there is *no single mechanism*, and differing interpretations in the litterature may not be contradictory, since they may apply to different conditions.

Above 1 μa the characteristic tightly adhering blue glow around the wire appears. This appears in point to plane corona, usually at the threshold here marked BP. The wire causes attenuation of luminosity by spread and requires larger currents to make it visible than in point to plane corona. Above onset the presence of an Ohms law[2] regime and the subsequent nearly expotential rise towards breakdown lead to the knee and the flat plateau noted in Fig. 5. These portions of the curves are of little interest except that where the source allowed of enough potential to look for breakdown streamers and breakdown, the streamers are indicated by vertical lines on the curves where they occur below 99 mm. The studies

were rarely carried to complete breakdown in order to preserve the wire for completion of the measurement sequence. It is interesting to note the presence of breakdown streamers as low as 27 mm. This indicates that contrary to expectation but in keeping with the work of L. H. Fisher and B. Bedersen,[15] as well as B. Gaenger, studying spark breakdown, streamers will occur at low pressures and gap lengths as long as overvoltages are adequate. At no time were *preonset streamers* observed in the Geiger counter regime. This was not surprising as the geometry in contrast to that in point to plane corona is not as convergent. In addition, preonset streamers are stimulated by water vapor contamination of the gas which was absent, and this circumstance with the low concentration of O_2 discourages streamer formation. The narrowness of the Geiger counter plateau follows from the clean Ni cathode cylinders, absence of limiting resistors, etc.

In the steady burst pulse corona regime, the well-known equations for the space charge limited current of Townsend and of Werner apply.[14] This, in terms of the Geiger counter threshold, can be used to calculate the mobility of the positive ions in N_2+1 percent O_2 using Eq. (6). For pressures of 715, 617, 345, 99, and 27 mm the values of the mobilities deduced by the Townsend equation to 20°C 760 mm are respectively 2.8, 2.8, 2.9, 2.5, and 2.9 cm²/v sec in contrast to the values of the order of 1.9 observed in *air*. This agreement is satisfactory in view of the rough nature of the theory.

Lowering the pressure lowers the threshold potential but not in proportion to the reduction in pressure. It is observed that the threshold potential, V_s, in all the studies with positive burst pulse corona, to which a threshold equation such as (3) is applicable, is not proportional to the gas pressure, p. That is, the surface field of the wire of radius r at threshold, X_{sr}, divided by the pressure p in mm is not a constant. Instead X_{sr}/p or V_s/p increases as pressure decreases.

Such behavior is to be expected on the basis of Eq. (3).[11,16] Neglect for the moment the effect of the factor βf as less important than the exponential term. Then for breakdown the value of the integral $\int_a^r \alpha dx$ must remain essentially constant as p is varied. For the coaxial cylindrical electrodes used in the study,

$$X=kV_r/x, \quad \text{with} \quad k=1/\log\frac{1.425}{0.00762}=0.19.$$

In consequence,

$$dx=\frac{kV_r}{X^2}dX,$$

$$\int_a^r \alpha dx=kV_r\int \alpha dX/X^2$$

$$=kV_r\int_0^{X_r/p}(\alpha/p)/(X/p)^2 d(X/p).$$

[15] L. H. Fisher and B. Bederson, Phys. Rev. **75**, 1615 (1949), **78**, 331 (1950). Also, G. A. Kachickas, Phys. Rev. **79**, 232 (1950). B. Gaenger, Archiv. f. Elektrotech. **39**, 508 (1949).

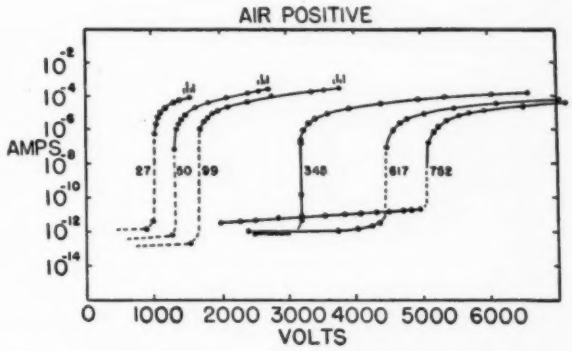


FIG. 6. Positive corona in air. Note the relatively slight slope of pre-threshold currents in air and the very sharp onsets. Note also the appearance of breakdown streamers indicated by vertical lines on curves for 99 mm and below.

With the electrodes used, $X_r=kV_r/r=25V_r$ so that

$$\int_a^r \alpha dx=0.19V_r\int_0^{(25V_r)/p}\frac{(\alpha/p)}{(X/p)^2}d(X/p). \qquad (7)$$

The form of (α/p) as a function of X/p is independent of p, so that the value of the integral depends upon V_r and p only as they appear outside the integral and in the upper limit of integration. Therefore, if p is reduced, V_r/p must increase as V_r is decreased in order for the integral to remain constant as required by Eq. (3) and as observed experimentally.

Consider now the effect of a variation of p on β. The absorption coefficient μ in $\beta=0.5\exp(-1.86\mu\bar{a})$ depends on pressure such that $\beta=0.5\exp(-1.86\mu_1 ap)$ where μ_1 is the coefficient at 1-mm pressure.[16] If μ_1 is large such that β is measurably affected by changes in $\mu_1 p$, decrease in pressure will materially increase β towards its limiting value of 0.5. Accordingly, if β increases the value of $\exp\int_a^r \alpha dx$ can decrease as pressure decreases to satisfy Eq. (3). Variation of β in this sense would compensate for the deficiencies in $\int \alpha dx$ and the relation would tend to yield a threshold V_s more nearly proportional to p. Again since X_{sr}/p increases as p decreases the average electron energy in the avalanches increases probably increasing f slightly. In the absence of accurate data on α/p for air the application of Eq. (7) over the range of conditions covered in this investigation is not possible. The calculations made indicate that the observations are primarily accounted for by the constancy of $\int \alpha dx$. Thus it can be concluded that β varies relatively little and that the value of μ for the burst pulse corona is more nearly 10 cm^{-1} than 100 cm^{-1} as indicated elsewhere.[11]

6. POSITIVE WIRE IN AIR

The results obtained with purified and dried room air are shown in Fig. 6. They exhibit nothing remarkable relative to the observations in N_2 with 1 percent O_2. The rather lower values of the preonset currents and their

[16] E. E. Dodd, Phys. Rev. **78**, 620 (1950).

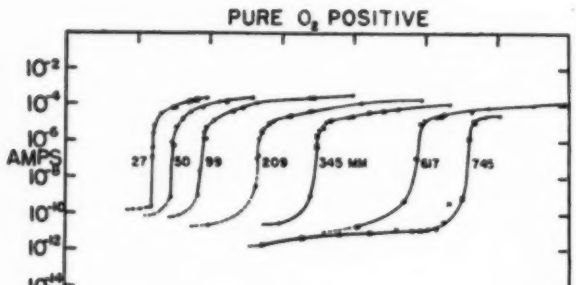


FIG. 7. Positive corona in pure oxygen. Note gradual rise below threshold with relatively shorter current jump at threshold. Note here that above 200 mm the corona consisted of streamers along the whole wire. While below 200 mm streamers and burst pulses appeared simultaneously with only burst pulses at 50 mm and below.

very slow increase with potential should be noted. This indicates that the O_2 has absorbed practically all of the photoelectrically active photons. It has perhaps rendered the Ni cylindrical cathode inactive and reduced the photoelectric emission from the cathode to such an extent that a general diffuse low order self-sustaining Townsend discharge is absent entirely. The curves indicate purely field intensified ionization currents imposed from outside until a corona threshold is reached. This is reached abruptly starting from 10^{-11} or less amp, depending on the radioactive triggering used, with a jump in current to 0.1 μa. The threshold is indicated by occasional burst pulses seen on the oscilloscope. These increase in frequency and duration with increasing potential until the bursts merge at the onset of the steady burst pulse corona which has no interruptions. The transition from intermittent to steady burst pulse corona was not clearly noted as with the point or the more dilute 1 percent O_2. In other words, the onset of steady corona was poorly defined, especially for 752- and 617-mm pressure. For 345-, 205-, 99-, and 27-mm pressures, the onset was more clearly defined and set in at 0.1 μa. Above 0.1 μa the corona current was irregular but did not show periods of extinction. This was probably caused by frequent and overlapping bursts at different points on the wire owing to the intense absorption leading to limited spread along the wire. In contrast to conditions at lower pressures, or with 1 percent O_2 in N_2, the photoelectric ionization in the gas was confined to a more limited region along the wire and to more rapid concentration of space charge, which extinguished the local bursts. With greater photon spread at lower pressures conditions reverted to those with diluted O_2. Even then the bursts pulses were both 2 to 4 times longer than for the point to plane corona in air and seemed to be of larger amplitude. The *increased amplitude* of the pulses probably reflects a more sensitive oscilloscope pick up with coaxial cylinders. As has recently been indicated by W. N. English,[17] the oscilloscope picks up only inductive signals from the charge liberated and its movement in the gap. With the large gaps in point to plane corona the signal is weak compared to the signals received by the coaxial Ni cylinder close to the wire.

From the steep rise of the curves of Fig. 6 at low pressure, it is seen that the Geiger counting plateaus in all these coronas are very narrow and ill defined. No preonset streamers were noted, although the irregularity of the burst pulse corona near onset at 752 and 617 mm indicates an intense absorption of photons. This action might be the forerunner of streamers at higher pressure, or with more O_2, as noted by Bandel[18] for point to plane corona. While such streamers were observed for some points or with moist air, it is unlikely that preonset streamers in dry air are favored by the coaxial geometry. The limitations of the potential source permitted carrying only the curves for 99, 50, and 27 mm to the regions where breakdown streamers could be observed. These were observed as shown by the vertical lines in Fig. 6. Data were taken with higher concentrations of O_2 but the curves were not plotted as they were of no significance.

7. POSITIVE WIRE PURE O_2

The results obtained with pure O_2 are summarized in the curves of Fig. 7. These data are again in marked contrast to air. The currents while starting at 10^{-12} ampere at atmospheric pressure begin to show an *appreciable slope well before the steep rise accompanying the initial threshold*, perhaps even more so than in pure N_2. This indicates some sort of ion multiplication preceding the threshold, which did not occur in air and has its steep rise nearly 500 v higher. There is no overshoot and curves are retraceable. This is generally characteristic of discharges whose secondary mechanism occurs *in the gas*. With the steep rise of current at 10^{-9} amp the oscilloscope showed *definite streamer pulses* and *no burst pulse corona* whatsoever. This must be interpreted as follows. The very high absorption of oxygen for the photo-ionizing photons prevents the spread of bursts down the wire and confines the discharges to a limited region. Thus, the cumulative avalanches triggered by photons have a value of the quantity β in Eq. (3) of such restricted scope that no bursts of large duration form. However, the field intensified currents produced by electron avalanches at the wire and possibly by the electrons liberated by *long wavelength photons* in O_2 from the Ni cylinder walls, are multiplied so as to give rather heavy currents, Townsend currents, following Eq. (2).† This could happen if in the absence

[17] W. N. English, Phys. Rev. **77**, 850 (1950).

[18] H. W. Bandel, paper in preparation, Department of Physics, University of California at Berkeley (1950).

† It appears that the steep pre-discharge currents are associated with avalanches triggered by photo-electrons from the Ni cylinder. These lead to avalanches noted in H_2 by Lauer. Since the discontinuous avalanches jumping from point to point on the cylinder surface do not choke of discharge as with burst pulses which spread uniformly over the wire pre-discharge currents

of N_2 and oxides of N, which influence the behavior in air, or N_2+O_2, the pure O_2 is strongly absorptive of photons capable of ionizing the gas, but is more transparent than air to longer wavelength photons capable of causing emission from Ni. Irrespective of such Townsend currents, when the field at the wire reaches the threshold for streamer formation, such as indicated by Eq. (4), then the current suddenly increases owing to the outward movement of streamers from the wire. The streamer onset of Eq. (4) is always higher than the burst pulse onset of Eq. (3). Thus, despite unfavorable geometry, streamers push outward into O_2 rather than spread the discharge laterally as burst pulses owing to the high absorption of O_2 for gas ionizing photons. The continuity of the predischarge current and the uniform spread of the streamers down the whole wire at once, as indicated by the current distribution in the cylinders, requires that *some transparency exist for photons capable of liberating electrons from the Ni cylinder walls*. The triggering observed could also be achieved by a secondary liberation of electrons by positive ions, but this would not be very effective in causing the observed diffusion of the discharge at higher pressures.

If this explanation is correct, then with an attenuation of O_2 molecule concentration equivalent to that for O_2 in air, by pressure reduction of pure O_2, i.e. at about 150-mm pressure of O_2, burst pulses should be observed along with the streamers. In fact, burst pulses were noted around 200-mm pressure in pure O_2.

As the potential in O_2 was raised above streamer threshold, the streamers became more frequent, yielding a proportionate increase in current. The increasing number of streamers seemed to favor appearances in groups as if the first streamer triggered a sequence of streamers which eventually locally fouled their section of the gap and extinguished the sequence. This observation could also have been a misinterpretation of statistical distributions by judgement made on appearances. When currents reached 5 μa the streamers had become quite constant in amplitude and occurred in groups of 3 to 10 streamers on the oscilloscope screen. The individual streamers were separated by about 280 μsec. The groups were separated by much larger intervals of random length. With further increase in current and potential, the group intervals shorten, and the groups are more closely spaced. A slight decrease in interval between streamers is noted as the potential increases. For example, at 617 mm and 9 μa the interval is 245 μsec, while at 20 μa, it is 172 μsec. What fixes the time between streamers at 280 μsec is hard to determine. If these were triggered by avalanches initiated by positive ion bombardment of the Ni cylinder, the assumption of normal ion mobilities and undistorted fields would give a time of crossing, and thus the interval as 1000 μsec instead of the 250 observed. This discrepancy is too large. It might favor the hypothesis of a photoelectric liberation from the Ni cylinder with a streamer build-up time in low fields of near 250 μsec as observed for sparks by L. H. Fisher and B. Bederson.[15] Separate groups of streamers are probably random discharges triggered off at different sections of the cylinder. The currents appeared properly distributed between central and guard rings. At 20 μa the corona first became *visible as a faint grayish glow along the wire*. In contrast to air this late visual appearance is not surprising for the bluish and purple glows of N_2 and oxides of N are much more luminous than the pale gray of O_2. Furthermore, while burst pulse corona concentrates ionization very close to the wire, the streamers diffuse it outward some 0.1 to 1 mm, thus decreasing visibility even more.

Increasing the current beyond 30 μa produces irregular streamer height and random time distribution. This appears to be caused by crowding into a single oscillograph pattern of impulses taking place over an extended length of wire. While individual pulses would be regular, superposition might weaken pulses by fluctuations of potential as well as superposition.

At 345 mm the threshold showed occasional random streamers. At 8 μa the heights of the streamers became uniform and repetition regular. At 13 μa groups with 160-μa separation between streamers appeared. Above 18 μa the streamers became so frequent and continuous that the current registered as continuous with fluctuations. Visual glow appeared at currents much higher than 12 μa.

At 209 mm the threshold current began at 10^{-9} amp rising to 0.1 μa and consisted of mixed streamers and some burst pulses. On increasing voltage, both streamers and burst pulses continued and large current fluctuations set in which reach a maximum at about 2 μa. Further increase in current results in space charge fouling by burst pulses which suppress streamers and lead to the onset of a steady burst pulse corona as observed in air, the discharge being general down the whole wire. The glow was visible only at 100 μa. At 99, 50, and 27 mm the threshold was marked by a jump in current from 10^{-9} amp to about 1 μa with only burst pulse corona. The low pressure so diffused the photo-electrons in the gas as to give only burst pulses and no streamers. The current was followed up to 200 μa at 99 mm with no streamers. The burst pulses spread down the wire as in air and at 50 and 27 mm lasted to 1 to 5 msec before extinction. This is to be expected with no current limiting resistor and beautifully indicates one of the causes leading to long dead times in unquenched Geiger counters. The glow became visible only at 200 μa to the central collector.

smoothly increase until avalanches form streamers. The current jump is then relatively small.

Surface-Wave Propagation Over a Coated Plane Conductor

STEPHEN S. ATTWOOD
Electrical Engineering Department, University of Michigan, Ann Arbor, Michigan
(Received August 9, 1950)

A *TM*-type wave is assumed propagated parallel to the surface of a plane conductor coated with a thin layer of dielectric. This wave is similar to the wave propagated over a dielectric-coated wire of circular cross section, discussed by Dr. Goubau in the references. Analysis of the cylindrical type wave requires the use of Hankel functions, while the wave over a plane surface may be described by simpler functions, namely, trigonometric functions in the dielectric layer and real exponential functions in the adjacent air region. The calculations thus are considerably simplified and the wave properties more readily observed.

Basic equations are developed for both dielectric layer and air regions, which indicate a criss-cross or multiply reflected wave in the dielectric and a unidirectional wave in the air. Equations are developed also for the electric flux line shapes, power propagated over the cross section, concentration of power flow in neighborhood of the film, attenuation due to conductor wall loss and dielectric film loss. Numerical calculations are given for five film thicknesses varying from 0.0001 to 0.01 meter and for five frequencies ranging from 3×10^8 to 3×10^{10} cycles per second.

INTRODUCTION

RECENT work by Dr. G. Goubau of the Signal Corps Engineering Laboratories, Fort Monmouth, has shown the practicability of generating and transmitting an electromagnetic wave of a type which heretofore has received scant attention. By coating a single conductor with a thin dielectric film, and transmitting a wave parallel to the conductor and film surfaces, it is possible to achieve a high concentration of energy in the neighborhood of the film, under conditions such that the losses may be materially less than is customarily found in wave guides. To obtain the special benefits obtainable from this type of single wire transmission line, film thicknesses of only fractions of a centimeter are required and the frequencies used will lie in the microwave range.

Dr. Goubau's original development involved the transmission of a *TM*-type wave in the neighborhood of a coated wire of circular cross section. The analysis of this cylindrical geometry requires the use of Hankel functions, which somewhat obscure the physical processes. The latter are illuminated more readily in the present article dealing with a plane conducting surface, since the functions required to describe a similar *TM*-type wave are trigonometric inside the film and real exponential in the surrounding air. The calculations thus are considerably simplified and the wave properties more readily observed.

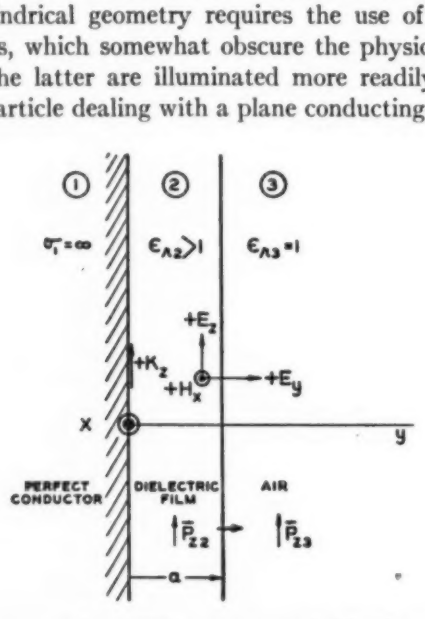


FIG. 1. Geometry for a *TM*-wave propagated in the *z*-direction.

THE FIELD EQUATIONS

Assumed is a perfect conductor ($\sigma_1=\infty$) with a plane surface, coated with a perfect dielectric (relative dielectric constant ϵ_{r2}) of thickness a, outside of which the dielectric (ϵ_{r3}) is vacuum (air). A *TM*-wave is supposed propagated in the z-direction through the two dielectrics parallel to the plane surface of the conductor. Figure 1 indicates the positive directions of the vector components of the wave, namely, H_x, E_y, E_z. For monochromatic propagation in the z-direction, we may remove the wave factor $\exp[j(\omega t-\beta z)]$ from Maxwell's equations. For the remainder, the three vectors are independent of x, and thus are functions only of y. Rationalized MKS units are used throughout.

With these limitations, and noting that $\partial/\partial z=-j\beta$, Maxwell's equations in rectangular coordinates reduce to:

$$(\partial E_y/\partial y)-j\beta E_z=0 \qquad (\mathrm{div}E=0), \quad (1)$$

$$(\partial E_z/\partial y)+j\beta E_y=-j\omega\mu H_x \qquad (\mathrm{curl}E=-j\omega\mu H), \quad (2)$$

$$\left.\begin{aligned} -j\beta H_x&=j\omega\epsilon E_y \quad (3)\\ -(\partial H_x/\partial y)&=j\omega\epsilon E_z \quad (4)\end{aligned}\right\} \qquad (\mathrm{curl}H=j\omega\epsilon E).$$

From (3)

$$H_x=-(\omega\epsilon/\beta)E_y. \quad (5)$$

Substitute H_x from (5) into (2), differentiate (1) to y, and combine. This gives the basic equation for E_z as

$$\partial^2E_z/\partial y^2=-(\omega^2\mu\epsilon-\beta^2)E_z. \quad (6)$$

Similarly,

$$\partial^2E_y/\partial y^2=-(\omega^2\mu\epsilon-\beta^2)E_y, \quad (7)$$

$$\partial^2H_z/\partial y^2=-(\omega^2\mu\epsilon-\beta^2)H_z. \quad (8)$$

For z-directed wave motion β must be real, and for y-directed wave motion $\omega^2\mu\epsilon>\beta^2$ in (6), (7), and (8).

Letting

$$k^2=\omega^2\mu\epsilon-\beta^2=(\omega/c)^2\mu_r\epsilon_r-\beta^2, \tag{9}$$

the equation for E_z, ignoring the wave factor, is

$$E_z=\tfrac{1}{2}(Ae^{+jky}+Be^{-jky}). \tag{10}$$

Similar solutions are obtainable for E_y and H_x.

In the dielectric layer, region 2, we have

$$E_{z2}=\tfrac{1}{2}(A_2e^{+jk_2y}+B_2e^{-jk_2y}). \tag{11}$$

At the conductor boundary, $y=0$, $E_{z2}=0$, whence $B_2=-A_2$, and

$$E_{z2}=\tfrac{1}{2}A_2(e^{+jk_2y}-e^{-jk_2y}) \tag{12}$$

$$=jA_2\sin(k_2y). \tag{13}$$

From (2) and (3),

$$E_{y2}=\tfrac{1}{2}A_2\frac{\beta}{k_2}(e^{+jk_2y}+e^{-jk_2y})$$

$$=A_2\frac{\beta}{k_2}\cos(k_2y) \tag{14}$$

and

$$H_{x2}=-\tfrac{1}{2}A_2\frac{\omega\epsilon_2}{k_2}(e^{+jk_2y}+e^{-jk_2y})$$

$$=-A_2\frac{\omega\epsilon_2}{k_2}\cos(k_2y). \tag{15}$$

In the outer dielectric, region 3, E_{z3} may be written in the form of (11), namely,

$$E_{z3}=[A_3'\exp(+jk_3'y)+B_3'\exp(-jk_3'y)]. \tag{16}$$

The phase constant $\beta=2\pi/\lambda$ in the wave factor must have the same value in both dielectric regions, since the wavelength in the z-direction of wave propagation must be the same on both sides of the boundary $y=a$. There are two choices for k_3'. A real value for k_3' indicates an unattenuated wave distributed sinusoidally in the y-direction. We shall, however, consider the other type of wave, for which k_3' is imaginary, which decreases as a real exponential in y but remains unattenuated in the z-direction. Experimental investigations have shown that this type of wave can be generated and propagated with properties of practical importance.[1]

Letting $k_3'=jk_3$ and retaining only the decreasing exponential, we observe further that A_3' must also be imaginary, since E_{z2} from (13) must equal E_{z3} at the boundary $y=a$. Letting $A_3'=jA_3$ and $B_3'=0$ we obtain

$$E_{z3}=jA_3e^{-k_3y}. \tag{17}$$

Further use of this boundary condition at $y=a$ gives

$$A_2\sin(k_2a)=A_3e^{-k_3a} \tag{18}$$

or

$$A_3=A_2e^{+k_3a}\sin(k_2a). \tag{19}$$

Then finally

$$E_{z3}=+jA_2e^{+k_3a}\sin(k_2a)e^{-k_3y} \tag{20}$$

and from (4) and (5)

$$H_{x3}=-A_2\frac{\omega\epsilon_3}{k_3}e^{+k_3a}\sin(k_2a)e^{-k_3y}, \tag{21}$$

$$E_{y3}=+A_2\frac{\beta}{k_3}e^{+k_3a}\sin(k_2a)e^{-k_3y}. \tag{22}$$

It is convenient to re-express the field components in terms of the surface current sheet on the conductor surface $y=0$. Let K_z equal the z-directed current per unit depth in the x-direction. Here the amplitude of K_z must equal the tangential component $(-H_{x2})$. From (15) we obtain $-A_2(\omega\epsilon_2/k_2)=K_z$. The field components, ignoring the wave factor $\exp[j(\omega t-\beta z)]$, then assume the forms:

$$E_{z2}=jK_z\frac{k_2}{\omega\epsilon_2}\sin(k_2y), \tag{23}$$

$$E_{y2}=K_z\frac{\beta}{\omega\epsilon_2}\cos(k_2y), \tag{24}$$

$$H_{x2}=-K_z\cos(k_2y) \tag{25}$$

and

$$E_{z3}=jK_z\frac{k_2}{\omega\epsilon_2}e^{+k_3a}\sin(k_2a)e^{-k_3y}, \tag{26}$$

$$E_{y3}=K_z\frac{k_2}{k_3}\frac{\beta}{\omega\epsilon_2}e^{+k_3a}\sin(k_2a)e^{-k_3y}, \tag{27}$$

$$H_{x3}=-K_z\frac{k_2}{k_3}\frac{\epsilon_3}{\epsilon_2}e^{+k_3a}\sin(k_2a)e^{-k_3y}. \tag{28}$$

So far the equations for both dielectrics are determined except for the numerical values of k_2, k_3, and β, which depend on the dielectric constants, the thickness of the dielectric film, and the frequency. The relations for k are given by (9).

$$k_2^2=(\omega/c)^2\mu_{r2}\epsilon_{r2}-\beta^2=\omega^2\mu_2\epsilon_2-\beta^2, \tag{29}$$

$$(k_3')^2=(\omega/c)^2\mu_{r3}\epsilon_{r3}-\beta^2=-k_3^2=\omega^2\mu_3\epsilon_3-\beta^2, \tag{30}$$

$$k_2^2+k_3^2=(\omega/c)^2(\mu_{r2}\epsilon_{r2}-\mu_{r3}\epsilon_{r3})=\omega^2(\mu_2\epsilon_2-\mu_3\epsilon_3). \tag{31}$$

At the boundary $y=a$, the tangential components of E and H are continuous, and hence the y-directed impedance is continuous, giving

$$Z_y=E_{z2}/H_{x2}]_{y=a}=E_{z3}/H_{z3}]_{y=a} \tag{32}$$

and

$$k_3=k_2(\epsilon_3/\epsilon_2)\tan(k_2a). \tag{33}$$

Combining (31) and (33) gives the equation for k_2 as

$$k_2^2[1+(\epsilon_3^2/\epsilon_2^2)\tan^2(k_2a)]=\omega^2(\mu_2\epsilon_2-\mu_3\epsilon_3). \tag{34}$$

[1] Georg Goubau, Signal Corps Engineering Laboratories; Surface wave transmission lines, Project No. 132A; Surface waves and their application to transmission lines, Technical Memorandum M-1260, February, 1950; also J. Appl. Phys. **21**, 1119 (1950).

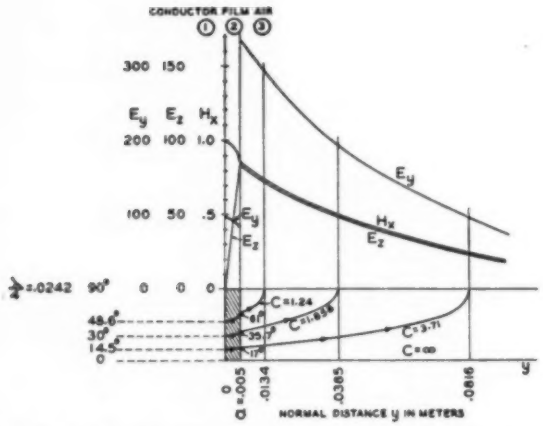


FIG. 2. Flux lines and field vectors for a particular case. Assumed $\sigma_1=\infty$, $\sigma_2=\sigma_3=0$, $a=0.005$ meter, $\epsilon_{r2}=4$, $\epsilon_{r3}=1$, $f=3\times10^9$ cycles per second.

Equation (34) may be solved for k_2, (33) for k_3, and (30) for β. From the latter

$$\beta^2=\omega^2\mu_3\epsilon_3+k_3^2=(\omega/v)^2 \tag{35}$$

in which v, the phase velocity in the z-direction for both dielectrics, is less, often slightly less, than c, the free space velocity. In general

$$v/c=\omega/\beta c. \tag{36}$$

MULTIPLE REFLECTIONS IN THE FILM

Omitting the time factor $e^{j\omega t}$ but restoring $e^{-j\beta z}$, (23), (24), (25) may be rewritten as

$$E_{z2}=\frac{K_z}{2}\frac{k_2}{\omega\epsilon_2}\left[e^{-j(-k_2y+\beta z)}-e^{-j(+k_2y+\beta z)}\right], \tag{37}$$

$$E_{y2}=\frac{K_z}{2}\frac{\beta}{\omega\epsilon_2}\left[e^{-j(-k_2y+\beta z)}+e^{-j(+k_2y+\beta z)}\right], \tag{38}$$

$$H_{x2}=-\frac{K_z}{2}\left[e^{-j(-k_2y+\beta z)}+e^{-j(+k_2y+\beta z)}\right]. \tag{39}$$

In each of these equations the first term represents a plane wave moving forward positively in z and negatively in y at an angle θ with the z-axis given by

$$\tan\theta=\frac{k_2}{\beta}=\left[(v/c)^2\mu_{r2}\epsilon_{r2}-1\right]^{\frac{1}{2}}. \tag{40}$$

The second term in each equation indicates a plane wave moving forward positively in z and positively in y at angle θ with the z-axis. Consequently, the forward wave in the z-direction in the film is a criss-cross, or multiply reflected, wave of two components, each progressing as a plane wave at angle θ with the z-axis, with phase velocity in the θ-direction of

$$v_d=\frac{1}{(\mu_2\epsilon_2)^{\frac{1}{2}}}=\frac{c}{(\mu_{r2}\epsilon_{r2})^{\frac{1}{2}}}. \tag{41}$$

The phase velocity in the z-direction, given by (36) and (29), is

$$v=\frac{\omega}{\beta}=\frac{v_d}{\cos\theta}=v_d\frac{(k_2^2+\beta^2)^{\frac{1}{2}}}{\beta}>v_d, \tag{42}$$

and the group velocity, the velocity of energy transport, is equal to

$$v_g=v_d\cos\theta=v_d\frac{\beta}{(k_2^2+\beta^2)^{\frac{1}{2}}}<v_d. \tag{43}$$

v and v_g are simply related through

$$vv_g=v_d^2=\frac{1}{\mu_2\epsilon_2}=\frac{c^2}{\mu_{r2}\epsilon_{r2}}. \tag{44}$$

The two plane wave components of E_{y2}, as given in (38), are directed instantaneously in the same direction; similarly for H_{x2} as given in (39). However, the two components of E_{z2} are instantaneously antiparallel, which is required to make the tangential value E_z at the conductor surface, $y=0$, always zero for a perfect conductor.

SHAPE OF ELECTRIC FLUX LINES

In a quarter-wavelength the electric flux lines emerge from the conductor surface charge, curve through the film, suffer refraction at the film-air surface, and curve through the air to return to the conductor in the next quarter-wavelength, as illustrated in the lower part of Fig. 2. The flux line shapes in region 3 (air) are developed from the basic equation

$$dy/E_{y3}=dz/E_{z3}, \tag{45}$$

in which E_{y3} and E_{z3} must be given in their instantaneous forms. Using the real parts of (26) and (27), including the factor $\exp[j(\omega t-\beta z)]$,

$$dz/dy=-(k_3/\beta)\tan(\omega t-\beta z). \tag{46}$$

For instant $t=0$,

$$dz/dy=+(k_3/\beta)\tan(\beta z), \tag{47}$$

which at once integrates to

$$k_3y=\ln\sin(\beta z)+\ln C$$

or

$$C\sin(\beta z)=\exp(+k_3y). \tag{48}$$

Various values of constant C determine the positions of various flux lines in air. The values indicated in Fig. 2 have been chosen to divide the field into equal increments (quarters) of flux and charge in a quarter wavelength.

The shapes of the flux lines in the film may be obtained similarly but are not given here since these portions of the lines are relatively unimportant for thin films.

POWER PROPAGATED

The time average power propagated in the z-direction, per meter in the x-direction, is equal to

$$\bar{P}_z=-\tfrac{1}{2}\int_{y=0}^{a}(E_{y2}H_{x2}^{*})_{\text{real}}dy -\tfrac{1}{2}\int_{y=a}^{\infty}(E_{y3}H_{x3}^{*})_{\text{real}}dy. \quad (49)$$

The first term gives the average power $\bar{P}_{z2}$ propagated in the film, the second term $\bar{P}_{z3}$ that which is propagated outside the film. H^* is the complex conjugate of H. From (24), (25), and (49)

$$\bar{P}_{z2}=+K_z^2\frac{1}{4k_2}\frac{\beta}{\omega\epsilon_2}[k_2a+\tfrac{1}{2}\sin(2k_2a)]. \quad (50)$$

Using (27), (28), and (49), with limits of $y=a$ to $y=\infty$, we get

$$\bar{P}_{z3}=+K_z^2\frac{1}{4}\frac{\epsilon_3}{\epsilon_2}\frac{k_2^2}{k_3^3}\frac{\beta}{\omega\epsilon_2}\sin^2(k_2a). \quad (51)$$

The last term, between the limits of $y=a$ to $y=b$, takes the form

$$\bar{P}_{z3}\Big]_a^b=+K_z^2\frac{1}{4}\frac{\epsilon_3}{\epsilon_2}\frac{k_2^2}{k_3^3}\frac{\beta}{\omega\epsilon_2} \times\sin^2(k_2a)\{1-\exp[-2k_3(b-a)]\}. \quad (52)$$

CONCENTRATION OF POWER DENSITY

An important feature of this type of wave is the fact that the power flow in the z-direction is highly concentrated in the film and the adjoining space. In the numerical illustrations given later the distance $y=b$ is calculated to include 90 percent of the power in the wave. (50) to (52) are useful for this purpose.

ATTENUATION FOR CONDUCTOR WALL LOSS

The preceding equations have been developed on the assumption of no loss. When the wall loss is small, as in this case, the corresponding attenuation factor is given by

$$\alpha_s=\frac{\bar{P}_s}{2\bar{P}_z}=\frac{\tfrac{1}{2}R_s[H_{x2}H_{x2}^{*}]_{y=0}}{2\bar{P}_z}. \quad (53)$$

in which $\bar{P}_s$ = time average power loss to the surface per unit length of line (in z) per unit width of line (in x) and R_s = surface resistance for skin effect per unit area = real part of the surface impedance $Z=(1+j)R_s=(\pi\mu_1/\sigma_1)^{\frac{1}{2}}(f)^{\frac{1}{2}}=2.609\times10^{-7}(f)^{\frac{1}{2}}$ for copper. ($\mu_1=4\pi\times10^{-7}$ and $\sigma_1=5.8\times10^{7}$ for copper.) Using (25), α_s proves to be equal to

$$\alpha_s=R_sK_z^2/(4\bar{P}_z). \quad (54)$$

The attenuation in db per meter is equal to $8.68\alpha_s$.

ATTENUATION FOR DIELECTRIC FILM LOSS

We have first to calculate the attenuation constant α_2 for the loss in the film, and second to calculate the attenuation constant α for the entire wave in both regions 2 and 3.

Maxwell's equations for an alternating steady-state wave in the lossy dielectric take the form

$$\text{curl}E_2=-j\omega\mu_2H_2, \quad (55)$$

$$\text{curl}H_2=(\sigma_2+j\omega\epsilon_2)E_2 =j\omega\left[\epsilon_2\left(1-j\frac{\sigma_2}{\omega\epsilon_2}\right)\right]E_2. \quad (56)$$

The latter indicates that the complex dielectric constant $\epsilon_2[1-j(\sigma_2/\omega\epsilon_2)]$ shall replace ϵ_2 when dielectric loss is to be taken into account.

Taking

$$\text{curl curl}E_2=\text{grad div}E_2-\nabla^2E_2$$

and noting that $\text{div}E_2=0$ in the space charge free dielectric, and substituting (56), we obtain

$$\nabla^2E_2=-\omega^2\mu_2\left[\epsilon_2\left(1-j\frac{\sigma_2}{\omega\epsilon_2}\right)\right]E_2 =\gamma^2E_2. \quad (57)$$

Here

$$\gamma_2=j\omega\left[\mu_2\epsilon_2\left(1-j\frac{\sigma_2}{\omega\epsilon_2}\right)\right]^{\frac{1}{2}} =\alpha_2+j\beta. \quad (58)$$

γ_2 is the complex propagation constant of the lossy dielectric and may be written in terms of α_2, the attenuation constant, and β, the phase constant for both dielectrics.

From (58)

$$\alpha_2=\omega\left(\frac{\mu_2\epsilon_2}{2}\left\{\left[1+\left(\frac{\sigma_2}{\omega\epsilon_2}\right)^2\right]^{\frac{1}{2}}-1\right\}\right)^{\frac{1}{2}} \quad (59)$$

and

$$\beta=\omega\left(\frac{\mu_2\epsilon_2}{2}\left\{\left[1+\left(\frac{\sigma_2}{\omega\epsilon_2}\right)^2\right]^{\frac{1}{2}}+1\right\}\right)^{\frac{1}{2}}. \quad (60)$$

For $(\sigma_2/\omega\epsilon_2)^2\ll1$, which holds for practical dielectrics in the microwave range,

$$\alpha_2\cong(\sigma_2/2)(\mu_2/\epsilon_2)^{\frac{1}{2}}, \quad (61)$$

$$\beta\cong\omega(\mu_2\epsilon_2)^{\frac{1}{2}}. \quad (62)$$

Combining (61) and (62) we obtain

$$\alpha_2\cong\tfrac{1}{2}(\omega/\beta)\mu_2\sigma_2 \quad (63)$$

$$\cong\tfrac{1}{2}(\mu_2c)(v/c)\sigma_2 \cong188.5(v/c)\sigma_2, \quad (64)$$

in which v is the phase velocity and c is the free space

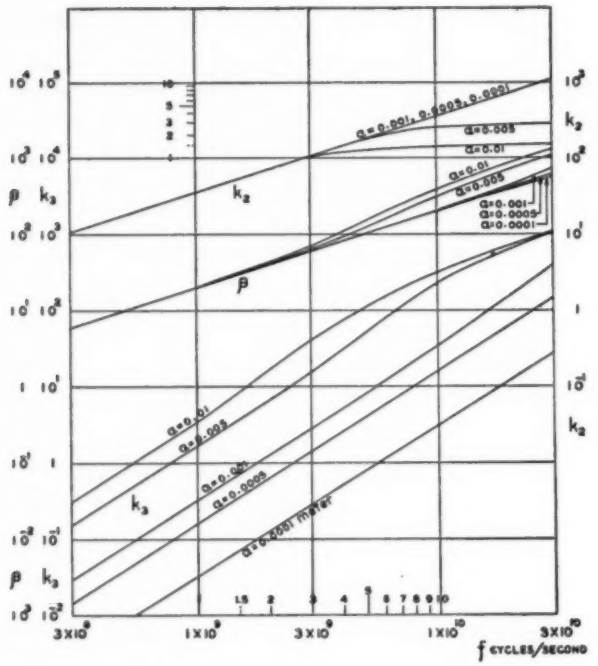


FIG. 3. Values of k_2, k_3, and β as functions of frequency.

velocity of the wave. The attenuation in the film in *db* per meter is equal to $8.68\alpha_2$.

The attenuation factor α for the dielectric loss for the entire wave in both dielectrics is given by

$$\alpha=\frac{\bar{P}_l}{2(\bar{P}_{x2}+\bar{P}_{x3})} \tag{65}$$

in which $\bar{P}_l=d\bar{P}/dz$ is the time average dielectric loss per meter (in z) per meter of width (in x). $\bar{P}_{x2}+\bar{P}_{x3}$ is the time average power propagated by the wave in both dielectrics per meter of width (in x).

$$\bar{P}_l=\frac{\sigma_2}{2}\int_{y=0}^{a}(E_{y2}E_{y2}^*+E_{z2}E_{z2}^*)dy. \tag{66}$$

Upon substituting (23) and (24)

$$\bar{P}_l=K_z^2\frac{\sigma_2}{4(\omega\epsilon_2)^2}\left[(\beta^2+k_2^2)a+\frac{\beta^2-k_2^2}{2k_2}\sin(2k_2a)\right]. \tag{67}$$

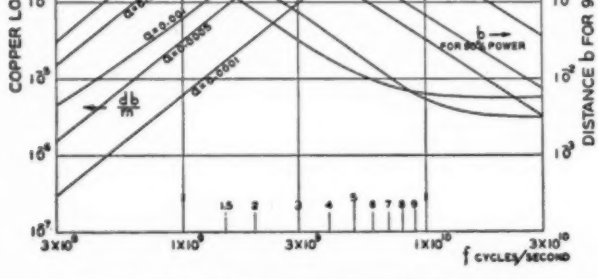


FIG. 4. Copper loss attenuation and 90 percent power concentration.

NUMERICAL CALCULATIONS

All calculations are based on a dielectric film whose relative dielectric constant $\epsilon_{r2}=4$ and an outer medium air with $\epsilon_{r3}=1$. Five film thicknesses are considered, varying from $a=0.0001$ to $a=0.01$ meter; and five frequencies, varying from 3×10^8 to 3×10^{10} cycles per second, have been used in the calculations.

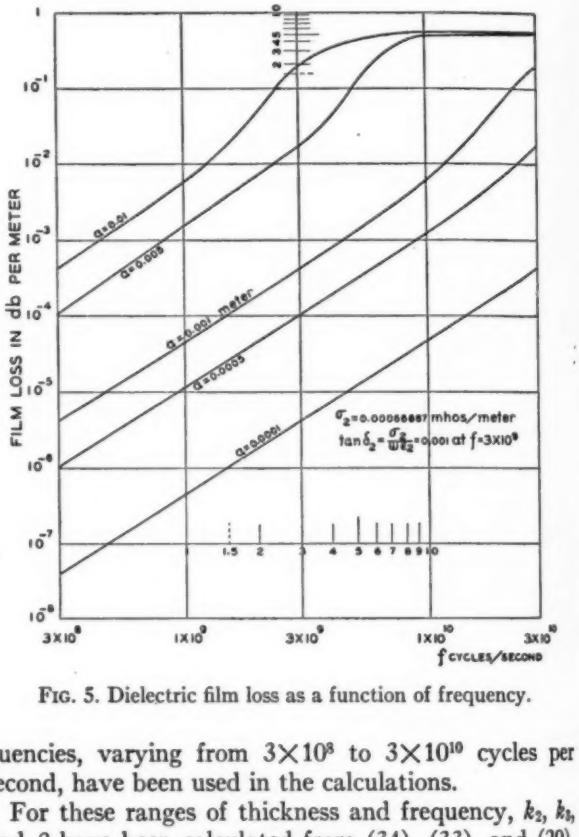


FIG. 5. Dielectric film loss as a function of frequency.

For these ranges of thickness and frequency, k_2, k_3, and β have been calculated from (34), (33), and (29), respectively, and are plotted in Fig. 3. These quantities all increase with frequency, and k_3 in particular increases also as the film thickness is increased, leading to higher y-directed attenuation and higher power-density concentration in and near the film. The distance b, for 90 percent power concentration, therefore decreases with frequency and for thicker films, as illustrated in Fig. 4 (see also (50) to (52)). The change in type of curvature occurs when the frequency is high enough so that most of the power is transmitted in the film.

Figure 4 also gives the copper conductor wall-loss attenuation as determined from (54). This quantity increases with frequency.

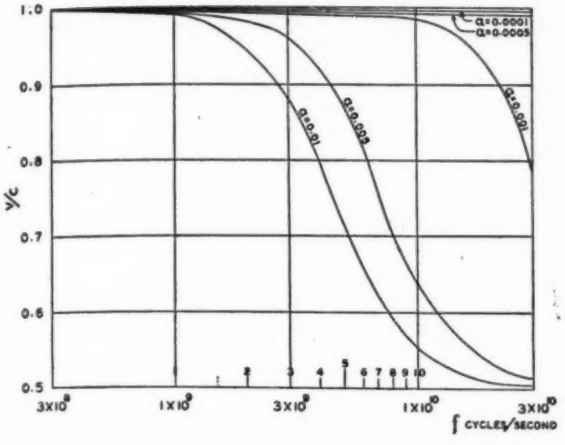


FIG. 6. Phase velocity ratio as a function of frequency.

For comparative purposes the attenuation for a copper-wall square wave guide is plotted as a broken line in Fig. 4. This plot is for a rectangular $TM_{1,1}$ wave, which is the one most nearly comparable to the wave type discussed in this paper. For each of the five frequencies calculated, the wave guide cross section has been adjusted to a cut-off frequency $f_c=f/\sqrt{3}$ which is the condition for minimum copper loss.

The dielectric loss in the film is calculated with the aid of (64) and is plotted in Fig. 5 for an assumed value of conductivity of $\sigma_2=0.00066667$, corresponding to a value of loss tangent of $\tan\delta_2=\sigma_2/(\omega\epsilon_2)=0.001$ at $f=3\times10^9$ cycles per second. The dielectric loss increases with both film thickness and frequency.

The phase velocity ratio v/c is plotted in Fig. 6. It falls only slightly below unity for low frequencies and thin films, but for thick films and high frequencies it approaches the asymptotic value of $1/(\epsilon_{r2})^{\frac{1}{2}}=0.5$. For the same ranges the group velocity ratio increases from $v_g/c=0.25$ to 0.5, in accordance with the relation $vv_g=c^2/(\mu_{r2}\epsilon_{r2})=c^2/4$.

The angles θ, at which the plane wave components of E_y, E_z, and H_x move with respect to the z axis, are given in Table I.

In the upper part of Fig. 2, E_z, E_y, and H_x are plotted as functions of y for $\epsilon_{r2}=4$, $a=0.005$ meter and $f=3\times10^9$ cycles per second. The quarter wavelength for these values is $\lambda/4=0.0242$ meter as compared with the free space quarter wavelength $\lambda_0/4=0.025$ meter, whence $v/c=0.0242/0.025=0.9689$ and the group velocity ratio $v_g/c=0.258$. The value of θ for this case is 58.93°.

TABLE I. Values of criss-cross angle θ in degrees.

f cycles / second	a in meters 0.0001	0.0005	0.001	0.005	0.01
3×10^8	60.	60.	60.	60.	59.97
1×10^9	60.	60.	60.	59.9	59.57
3×10^9	60.	59.99	59.97	58.93	53.55
1×10^{10}	60.	59.9	59.57	38.85	20.42
3×10^{10}	59.97	58.93	53.55	13.8	7.01

CONCLUSION

The objectives of this article are to explore the characteristics and study the magnitudes of a wave type which, until recently, has received no attention. The calculated values of low attenuation and high concentration of power in the neighborhood of the film suggest the need for exploring the possibility of utilizing this type of wave in practical applications.

Consideration of the equations and the calculations indicate that the latter do not represent optimum conditions. Clearly, improved results can be obtained by using a film with a higher value of dielectric constant and a lower value of conductivity. These changes would improve the concentration of power and lower the dielectric loss. The power concentration can also be increased by using geometries other than the plane—for instance, the cylindrical geometry used by Dr. Goubau.

Though not explored in this article, it would seem that the principles given might be used to improve wave-guide design by concentrating the power near the walls. Such a design might lower the guide attenuation and permit use of cross-section sizes relatively independent of the frequency used.

The Mathematical Basis of the Interpretation of Tracer Experiments in Closed Steady-State Systems*

C. W. SHEPPARD AND A. S. HOUSEHOLDER
Biology Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee

(Received October 6, 1950)

The basic equation for the mixing of radioactive isotopes is $\mathbf{dR/dt}=I\ \mathbf{grad}\ a$, where $\mathbf{dR/dt}$ is the flux of radioactivity; a, the relative specific activity; and I, the interfusion constant. The equation is applied to systems of compartments and some electrical analogies are presented. As special cases of the unconstrained systems treated in an earlier paper two constrained systems are discussed in which only a single radioactive isotope is required for the study of transport phenomena. One of these (mammillary system) consists of a central compartment communicating with a number of peripheral compartments. In the other (catenary system) the compartments are arranged chainwise. Equations are derived for the n-compartment mammillary system relating the time dependent specific activities of the compartments to the exchange rates between the peripheral and central compartments and to their content of substance being exchanged. In analyzing the effect of lumping of a series of compartments into a single equivalent one the case where the exchange rates cluster about a mean is discussed in some detail. In the catenary system the greater mathematical difficulties are indicated including the fact that the results depend on the order of the compartments. It is pointed out that in mammalian physiological experiments observations confined to only one compartment such as the circulating blood are insufficient for a complete analysis.

INTRODUCTION

NUMEROUS examples of the use of equations in interpreting isotope experiments have appeared in the literature during the past few years. A recent review by Solomon[1] considers some of the special equations encountered in physiological and biochemical studies with isotopes. Because of the growing importance of the isotope as a biological research tool it is felt that a more fundamental treatment of the underlying mathematical principles is in order. Such a discussion at this time should help to systematize the growing notation and terminology in this field, and to constrain such theories within the convenient framework of a concise logical system of maximum generality but possessing the minimum essential elements and postulates.

The practically valuable but theoretically trivial use of the radioactive isotope as a sensitive method of determining trace concentrations of chemical elements will be excluded from the present discussion. The determination of quantities of material by isotope dilution has already been well reviewed.[2] There remains then the study of the process of mixing of tagged and untagged species of a substance S in a system following the local introduction of an aliquot of isotopically labeled material. The concentration of S may vary spatially throughout the system. Maximum generality would also admit variation of concentration with time at any point. Since this situation is considerably more complicated its inclusion must be postponed except for special instances. We are thus concerned with mixing by exchange. We will also limit the discussion to closed systems where S does not flow in or out through the enclosing boundaries. A general discussion of open systems is beyond the scope of the present communication.

THE MIXING OF ISOTOPES BY EXCHANGE—THE FICK LAW ANALOGY

The mixing of molecules of a solute with a solvent or the general process of achieving uniform concentration is expressed by Fick's Law

$$\mathbf{dS/dt}=-D\ \mathbf{grad}\ C, \tag{1}$$

where dS/dt is the amount of S flowing per unit area per second, D the diffusion constant (assumed isotropic), and C the concentration of S. The mixing of isotopes must not be confused with this process since it is the mixing of labeled and unlabeled species of S with one another in the approach to uniform isotopic composition. This can occur in a system that is maintained by external forces in a steady state completely out of thermodynamic equilibrium. A convenient term for it is interfusion. An analogous equation is derived in Appendix A, namely,

$$\mathbf{dS^*/dt}=-I\ \mathbf{grad}\ A. \tag{2}$$

The vector quantity $\mathbf{dS^*/dt}$ represents the rate of flow of a given labeled species of S per unit area. The scalar point function I is called the interfusion constant, which we will assume has the same value for both tagged and untagged species of S. The quantity A represents the fractional amount of S that is tagged and will be called the *absolute specific activity*.

An alternative expression in the case of radioactive isotopes is

$$\mathbf{dR/dt}=-I\ \mathbf{grad}\ a, \tag{3}$$

where $\mathbf{dR/dt}$ is the rate of flow of radioactivity per unit area and a the radioactivity per unit amount of S called the *relative specific activity*.

* Word done under Contract No. W-7405-Eng-26 for the AEC.
[1] A. K. Solomon, J. Clin. Invest. **28**, 1297 (1949).
[2] Gest, Kamen, and Reiner, Arch. Biochem. **12**, 273 (1947).

The analogy between Eq. (1) and Eqs. (2) and (3) is convenient since the solution of equations encountered in isotope problems may occasionally be found in the literature on diffusion.[3] The equations are of the same form as the classical equation of Fourier for the conduction of heat. Other examples such as the flow of compressible fluids in porous media can be cited.[4] A discussion of the use of electrical analogies will be given subsequently. The present formulation is chosen because it emphasizes the basic similarity of many problems in mathematical physics whose partial differential equations are solved by familiar standard methods.[5]

COMPARTMENTS

Although the use of isotopes in the study of processes in systems of uniformly variable composition may later be developed, the principal successes up to the present are in systems of compartments. The analogous case of compartments in the theory of heat conduction would be the transfer of heat between copper blocks at different temperatures separated by layers of material of low thermal conductivity. In isotope studies compartments may be regions of space in which A is uniform such as erythrocytes and plasma *in vitro*[6] or states of uniform chemical composition such as copper ions and a copper chelate compound.[7] There is some mental advantage perhaps in visualizing compartments as regions of space separated by boundaries and as such they will be represented here, although geometrically a lattice point description is more concise. We recognize the frequent difficulty, in biological experiments, of the isolation of compartments and proof of the constancy of A.

Multicompartment Systems

The most general approach to the study of the motion of labeled materials in a system of compartments is to consider first the situation where there are any number n of compartments. The system is not constrained and material can move from any compartment to any other. From such a treatment it is thus possible to derive certain general conclusions, and to obtain equations which can then be reduced to specific cases. This was done in an earlier communication.[8] The basic derivation will be repeated here in a more concise form.

Consider a system of n compartments (Fig. 1) containing n individual species of a substance S, these being uniformly mixed and individually identifiable, but otherwise behaving identically. In compartment j

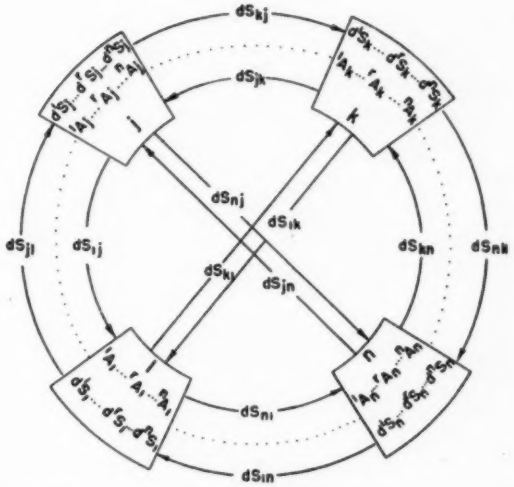


FIG. 1. The unconstrained multicompartment system. Compartments are numbered from 1 to n. Compartment k contains a relative amount ${}^{r}A_k$ of the rth species of S. The amount of S going from compartment j to compartment k is dS_{kj}.

there will be ${}^{r}S_j$ units of species r and its fractional amount will be ${}^{r}A_j$. The A's may be expressed as the elements of a matrix A, the r being the row index and the j the index of columns. If transfers occur such that dS_{jk} units are delivered from compartment k to compartment j the dS_{jk} may be determined in a single compartment by finding the ${}^{r}A_k$ in all compartments and the changes of all species $d^{r}S_j$ in compartment j only. The dS_{jj} is the amount remaining in j.

The equation relating the aforementioned quantities is obtained as follows. In a given compartment (say, j) the amount of a given species leaving it and going to compartment k is ${}^{r}A_j dS_{kj}$. The net change in compartment j is thus:

$$d^{r}S_j=\sum_{k=1}^{k=n}({}^{r}A_k dS_{jk}-{}^{r}A_j dS_{kj}). \tag{4}$$

We define the reciprocal matrix $\|\alpha_{jr}\|$ such that

$$\sum_{r}\alpha_{jr}\,{}^{r}A_k=\delta_{jk},$$

where

$$\delta_{jk}=0 \quad \text{if} \quad j\neq k \quad \text{and} \quad \delta_{jj}=1.$$

Multiply both sides of Eq. (4) by α_{ir} $(i\neq j)$ and sum over r giving

$$\begin{aligned}\sum_{r}\alpha_{ir}d^{r}S_j&=\sum_{r}\alpha_{ir}\sum_{k}({}^{r}A_k dS_{jk}-{}^{r}A_j dS_{kj})\\&=\sum_{k}(\delta_{ik}dS_{jk}-\delta_{ij}dS_{kj})\\&=dS_{ji}.\end{aligned} \tag{5}$$

But from the known properties of the reciprocal matrix[9]

$$\alpha_{ir}=C_{ir}/A,$$

C_{ir} being the cofactor of ${}^{r}A_i$ in the determinant A. In

[3] M. H. Jacobs, Ergeb. Biol. **12**, 1 (1935).
[4] M. Muskat, *The Flow of Homogeneous Fluids through Porous Media* (McGraw-Hill Book Company, Inc., New York, 1937).
[5] H. S. Carslaw, *Introduction to the Mathematical Theory of Conduction of Heat in Solids* (Macmillan Company, Ltd., London, 1921).
[6] C. W. Sheppard and W. R. Martin, J. Gen. Physiol. **33**, 703–722 (1950).
[7] R. B. Duffield and M. Calvin, J. Am. Chem. Soc. **68**, 557 (1946).
[8] C. W. Sheppard, J. Appl. Phys. **19**, 70 (1948).
[9] O. Veblen, *Invariants of Quadratic Differential Forms* (Cambridge University Press, New York, 1933).

the expanded form,

$$dS_{jk}=\frac{\begin{vmatrix} {}^1A_1 & \cdots & {}^1A_{k-1} & d^1S_j & {}^1A_{k+1} & \cdots & {}^1A_n \\ \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\ {}^nA_1 & \cdots & {}^nA_{k-1} & d^nS_j & {}^nA_{k+1} & \cdots & {}^nA_n \end{vmatrix}}{\begin{vmatrix} {}^1A_1 & \cdots & {}^1A_n \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ {}^nA_1 & \cdots & {}^nA_n \end{vmatrix}}. \quad (6a)$$

This expression is most conveniently used in mass spectrometer studies where ${}^rA_k$ is the relative abundance of isotope r in compartment k. In using radioactive isotopes the quantities ${}^rR_k$ and ${}^ra_k$ are measured. These are, respectively, the radioactivity per cc and specific activity of isotope r in compartment k. (It must be possible to distinguish separate radioactivities one from the other.) In this case it has been shown[8] that

$$dS_{jk}=\frac{\begin{vmatrix} {}^1a_1 & \cdots & {}^1a_{k-1} & V_jd^1R_j & {}^1a_{k+1} & \cdots & {}^1a_n \\ \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\ 1 & \cdots & 1 & V_jdC_j & 1 & \cdots & 1 \end{vmatrix}}{\begin{vmatrix} {}^1a_1 & \cdots & {}^1a_n \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ 1 & \cdots & 1 \end{vmatrix}} \quad (6b)$$

C_j being the concentration of S in compartment j and V_j the volume of the compartment. Since the denominator must not vanish, the relative amounts of the different species of S must be different in different compartments. The measurements required in Eq. (6a) are determinations of the relative amounts of n species of S in all compartments and determinations of the changes of each species in the receiving compartment only.

In experiments with radioactive isotopes, $n-1$ different types of radioactivities must be measured in each of the n compartments. In general, therefore, access to all compartments is required.

The problem of determining exchanges in an unrestricted multicompartment system in one experiment has considerable practical limitation due to the likelihood that the contents of the compartments will not be uniformly mixed. Also, considerable labor is involved in obtaining a proper record of each ${}^rA_k$ and d^rS_j during the experimental period.

The Two-Compartment Case

As an illustration of how the general case of n compartments is particularized we consider the two-compartment system in which opposite and nonequal transfers can be determined using a single radioactive label of relative specific activity a_1 or a_2. Here, inserting $n=2$ into Eq. (6b) and simplifying,

$$dS_{12}=S_1da_1/(a_2-a_1), \quad dS_{21}=S_2da_2/(a_1-a_2). \quad (7)$$

The subscripts relate the quantities to their respective compartments.

We consider for the moment the converse task of predicting the variation of the a's where dS_{12} or dS_{21} is known. One instance will serve as an illustration. In the case where S is exchanged at equal and opposite rates ρ, the equations are easily integrated inserting the boundary conditions $[t=0,\ a_1=a_1(0),\ a_2=0]$ obtaining the familiar expressions,

$$\left.\begin{aligned} a_1/a_1(0)&=\frac{S_1+S_2e^{-\rho t(1/S_1+1/S_2)}}{S_1+S_2} \\ a_2/a_1(0)&=\frac{S_1(1-e^{-\rho t(1/S_1+1/S_2)})}{S_1+S_2} \end{aligned}\right\}. \quad (8)$$

Since

$$\log[(a_1-a_2)/a_1(0)]=\rho t(1/S_1+1/S_2)=0.69t/T_{\frac{1}{2}}, \quad (9)$$

a plot of $(a_1-a_2)/a_1(0)$ on semilogarithmic coordinate paper will yield a straight line whose slope determines the half-value time $T_{\frac{1}{2}}$ and thus the exchange rate since

$$T_{\frac{1}{2}}=0.69S_1S_2/\rho(S_1+S_2).$$

It is of interest that the equations are valid even if the volumes of the compartments are changing during the experiment.

In principle the criterion of uniform mixing must always hold. Nevertheless, initially when the radioactivity of compartment 2 is very small, uniform mixing is required only in compartment 1. At this point

$$dS_{21}/dt=[-S_1da_1/a_1dt]_{t=0}. \quad (10)$$

If compartment 2 contains several subcompartments exchanging at different rates with compartment 1 this equation also applies, but dS_{21}/dt will now be the sum for the several subcompartments.

In a system such as the erythrocytes and plasma *in vivo* in a mammalian experiment, if a proper record of the a's and S's is obtained throughout the experimental period Eqs. (7) still apply in determining the dS_{12} and dS_{21}. This results from the restriction that the cellular compartment exchanges only with the plasma. Since the plasma exchanges with the extravascular system Eqs. (8) cannot be used and pure exponential changes cannot be expected. This point was correctly appreciated by Dean, *et al.*,[10] but the results were complicated by difficulties in obtaining a good record of the observed variables.

[10] Dean, Noonan, Haege, and Fenn, J. Gen. Physiol. **24**, 353 (1940).

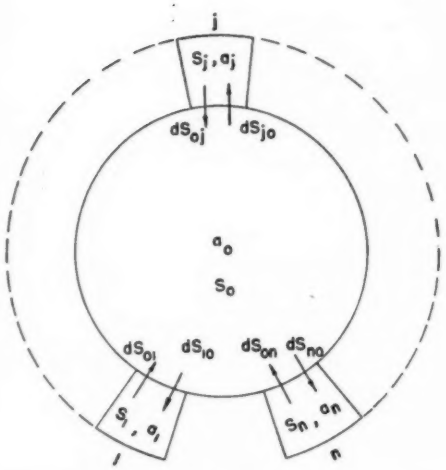


FIG. 2. The $n+1$ compartment mammillary system. The central compartment communicates with n peripheral ones.

CONSTRAINED MULTICOMPARTMENT SYSTEMS

In many instances a multicompartment system is constrained so that the exchange between many pairs of compartments is zero. This will not invalidate use of Eqs. (6) for the determination of the dS_{jk}. However, if any dS_{jk}'s are known (particularly if they are zero) the number of necessary observations is reduced and the equations may be correspondingly simplified. By transposing all the terms containing known dS_{jk}'s to the left side of Eqs. (4) it is evident that for every such known quantity one less equation is required and correspondingly one less $d'S_j$ need be determined. In the case of an n compartment system requiring n isotopes all equations containing measurements involving a given species of S may be removed for every n known constraints in dS_{jk}. Thus, one less species of S is required for the determination of the remaining dS_{jk}. Two particular systems are of interest in this regard. The first (Fig. 2) will be called a "mammillary" system where a central compartment communicates with a series of peripheral ones. This system may perhaps be used with some reservations to approximate the plasma of mammalian blood *in vivo*. The second situation is the "catenary" system where compartments are arranged chainwise. This system (Fig. 3) arises most commonly in the study of metabolic phenomena. In both systems because of the nature of the constraints only two species of S are required for the determination of the dS_{jk}. When a radioactive label is used this usually involves a determination of radioactivity and of total S.

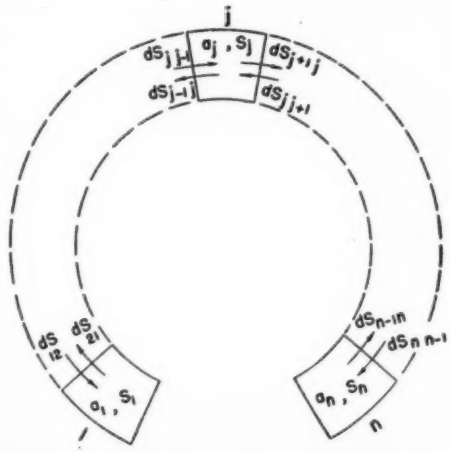


FIG. 3. The n compartment catenary system. The n compartments communicate chainwise.

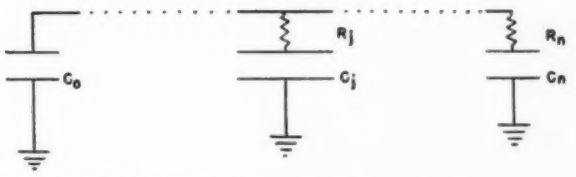


FIG. 4. Electrical analogy for the mammillary system.

Electrical Analogies

In the case of exchange in closed steady-state systems electrical analogies provide some convenience in understanding the processes which occur. They also suggest the possibility of using analog computers to reproduce the behavior of multicompartment systems containing isotopic tracers. As an illustration the electrical network in Fig. 4 is analogous to the mammillary system if we make the following substitutions:

1. The total radioactivity of compartment j becomes the charge on condenser j.
2. The specific activity of compartment j becomes the voltage across condenser j.
3. The total amount of S in the compartment becomes the capacitance C_j of the condenser.
4. The exchange rate between j and the central compartment becomes the conductance $1/R_j$ of the resistor j.

A similar analogy (Fig. 5) exists for exchange processes in the catenary system.

THE MAMMILLARY SYSTEM

By considering the general case of n peripheral compartments expressions may be obtained which can be particularized for any value of n. Only the exchange of S under steady state conditions will be considered. For

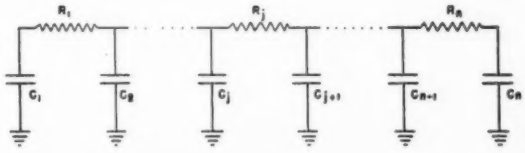


FIG. 5. Electrical analogy for the catenary system. This applies only to exchange processes and does not include turnover problems.

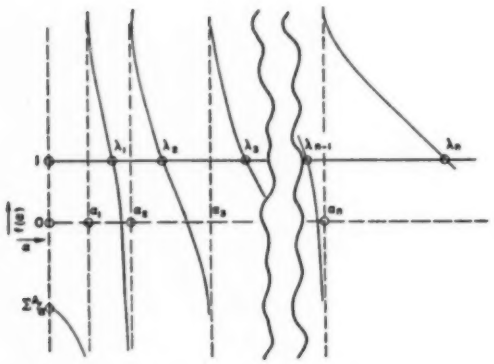


FIG. 6. Location of the roots of $g(\lambda)=0$. The λ_i are found at the intersection of the curves

$$f_1(\alpha)=\sum_1^n \frac{A_i}{\lambda-\alpha_i} \quad \text{and} \quad f_2(\alpha)=1.$$

The first $n-1$ λ_i lie between the asymptotes α_1 to α_n of $f_1(\alpha)$, and λ_n lies beyond α_n.

purposes of simplification we adopt the following notation:

ρ_j = exchange rate between the jth compartment and the central one.
S_0 = amount of S in the central compartment.
S_j = amount of S in the jth peripheral compartment.
S = total amount of S in the system.
$A_j=\rho_j/S_0$ (not related to the A_j above).
$\alpha_j=\rho_j/S_j$.
a_0 = specific activity in the central compartment.
$a_0(0)$ = initial value of a_0.
a_j = specific activity in the jth compartment.
λ_i = ith exponential constant in the equations $a(t)$.
x_i = coefficient of the ith exponential term in $a_0(t)$.

These quantities are related through the differential equations,

$$\left.\begin{aligned} da_0/dt&=\sum_k (a_k-a_0)A_k \\ da_j/dt&=(a_0-a_j)\alpha_j \quad (j=1, 2, \cdots, n)\end{aligned}\right\} \tag{11}$$

which show that the exchange rate between a given compartment and the central one can be determined from observations in these two compartments only.

To predict the behavior of the system under steady state conditions knowing the exchange rates and the values of S_0 and S_j requires a solution of Eqs. (11) which is obtained by familiar methods[11] in the form,

$$\begin{aligned} a_0/a_0(0)&=x_0+\sum_i x_i \exp(-\lambda_i t) \\ a_j/a_0(0)&=x_0/\alpha_j+\sum_i (\alpha_j-\lambda_i)^{-1}x_i \exp(-\lambda_i t).\end{aligned} \tag{12}$$

[11] T. von Karman and M. A. Biot, *Mathematical Methods in Engineering* (McGraw-Hill Book Company, Inc., New York, 1940).

The λ_j are the roots of

$$\Delta(\lambda)=\begin{vmatrix} \lambda-\sum A_i & A_1 & A_2 & \cdots & A_n \\ \alpha_1 & \lambda-\alpha_1 & 0 & \cdots & 0 \\ \alpha_2 & 0 & \lambda-\alpha_2 & \cdots & 0 \\ \cdot & \cdot & & & \\ \cdot & \cdot & & & \\ \cdot & \cdot & & & \\ \alpha_n & 0 & 0 & \cdots & \lambda-\alpha_n \end{vmatrix} = \lambda\phi(\lambda)g(\lambda)=0 \tag{13}$$

where

$$\phi(\lambda)=\prod_j (\lambda-\alpha_j) \tag{14}$$

and

$$g(\lambda)=1-\sum_i [A_i/(\lambda-\alpha_i)]. \tag{15}$$

In addition to the trivial root $\lambda_0=0$ there are n others obtainable from the relation $g(\lambda)=0$. Figure 6 shows that

$$\alpha_1<\lambda_1<\alpha_2<\lambda_2<\cdots<\alpha_n<\lambda_n.$$

For the boundary conditions, $t=0$, $a_0=a_0(0)$, $a_j=0$

$$\begin{aligned} x_0+\sum x_i&=1, \\ x_0/\alpha_j+\sum_i x_i/(\alpha_j-\lambda_i)&=0. \end{aligned} \tag{16}$$

The x's can be defined as functions of the system parameters A_j and α_j by an artifice as follows: Consider the quotient $\phi(\lambda)/\Delta(\lambda)$ whose numerator is a polynomial of degree n and whose denominator is a polynomial of degree $n+1$. The zeros of the denominator are $\lambda_0=0$ and the λ_i. Hence the quotient can be resolved into partial fractions in the form,

$$\phi(\lambda)/\Delta(\lambda)\equiv C_0/\lambda+\sum C_i/(\lambda-\lambda_i), \tag{17}$$

where the C's are constant. By virtue of Eq. (13) this can also be written

$$[\lambda g(\lambda)]^{-1}\equiv C_0/\lambda+\sum C_i/(\lambda-\lambda_i). \tag{18}$$

These are algebraic identities and must be satisfied for any value of λ. In particular if we set $\lambda=\alpha_j$, then since $\phi(\alpha_i)=0$, we have

$$0=C_0/\alpha_j+\sum C_i/(\alpha_j-\lambda_i).$$

These equations for $j=1, 2, \cdots, n$, are the same as the last n of Eqs. (16) with the C's replacing the x's. Also (18) can be written

$$[g(\lambda)]^{-1}\equiv C_0+\lambda\sum C_i/(\lambda-\lambda_i),$$

and as $\lambda\to\infty$ we have in the limit

$$1=C_0+\sum C_i. \tag{19}$$

Replacing C_i with x_i in this equation we obtain the first Eq. (16).

Since the C's and the x's satisfy the same equations in every respect and since these equations are sufficient to determine them uniquely, they are alternative representatives of the same quantity. The x's are thus

defined by

$$[\lambda g(\lambda)]^{-1} \equiv x_0/\lambda + \sum_i x_i/(\lambda-\lambda_i). \qquad (20)$$

To determine the x's from this defining relation we expand in partial fractions using the usual standard identity. The equation is multiplied through by λ and taking the limit $\lambda \to \infty$ thus,

$$[\lambda g(\lambda)]^{-1} = \phi(\lambda)/\Delta(\lambda) \equiv \sum_i \phi(\lambda_i)/\Delta'(\lambda_i)(\lambda-\lambda_i), \quad (21)$$

and

$$x_i = \phi(\lambda_i)/\Delta'(\lambda_i).$$

By the definition of the λ_i as the roots of $\Delta(\lambda)$

$$\Delta(\lambda) = \prod_j (\lambda-\lambda_j),$$

and

$$\Delta'(\lambda_i) = \prod_{i \neq j} (\lambda_i-\lambda_j),$$

so

$$x_i = \prod_k (\lambda_i-\alpha_k) / \prod_{j \neq i} (\lambda_j-\lambda_i)$$

(including λ_0 in the denominator), (22)

since

$$\lambda_0 = 0$$

$$x_0 = \prod_i \alpha_i/\lambda_i = 1/g(0) = S_0/\sum_0^n S_j = S_0/S.$$

The Three-Compartment Case

Appendix B illustrates the application of the result to the three-compartment case. Equations for such systems have been discussed in the literature, the mammillary case by Gellhorn, Merrell, and Rankin[12] (for the case $\alpha_1 \gg \alpha_2$) and the catenary case by Cohn and Brues.[13] In this particular instance the two systems are identical and the solutions will differ only in the x's which are determined from different boundary conditions. The exponents are obtained by solving Eq. (13) for $n=2$, and after appropriate changes in notation yields their results. Since this equation is essentially of degree $(n-1)$, explicit expressions for the λ_i cannot be obtained for systems of more than five compartments. Because of the increasing complexity of the situation the explicit treatment of systems of more than three compartments is scarcely warranted. The utility of the general treatment will appear in special instances where approximations may be made.

The Effect of Lumping Two Peripheral Compartments

One such approximation occurs when $\alpha_1 \to \alpha_2$. Here as an illustration (see Appendix) the effect of lumping of the two peripheral compartments into a single

[12] Gellhorn, Merrell, and Rankin, Am. J. Physiol. **142**, 407 (1944).

[13] W. E. Cohn and A. M. Brues, J. Gen. Physiol. **28**, 449 (1945).

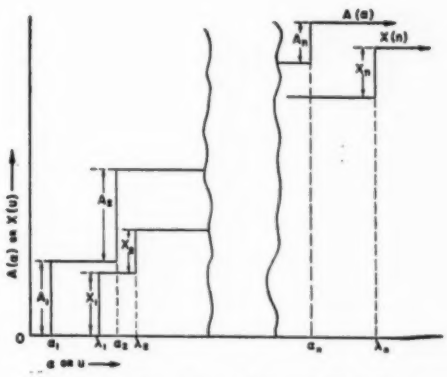


FIG. 7. The step functions $A(\alpha)$ and $X(u)$ (see text).

average one will be investigated. In this approximation

$$\lambda_1 \to (\alpha_1+\alpha_2)/2 = \alpha$$
$$\lambda_2 \to \alpha + A_1 + A_2 = \alpha + \sum A, \qquad (23)$$

$$x_1 \to \Delta\alpha^2/4\alpha\sum A$$
$$x_2 \to \sum A/(\sum A+\alpha). \qquad (24)$$

The $\sum A$ is the sum of the exchange rates relative to the contents of the central compartment and α is the arithmetic mean of the exchange rates relative to the contents of the peripheral compartments. If the spread $\Delta\alpha$ is 20 percent of either $\sum A$ or α, the x_1 will be <0.01. Thus the behavior of the system is practically identical to that of an equivalent two-compartment system. It is of considerable interest that the result is so insensitive to the spread of the α's.

CONTINUOUS DISTRIBUTION OF PERIPHERAL COMPARTMENTS

In some biological experiments such as sodium or potassium exchange in human erythrocytes *in vitro* a central compartment (the plasma) communicates with a very large number of peripheral ones (the cells). Here, although the properties of the cells will be variable, this variation is ignored and the total group of cellular compartments is customarily lumped into a single average compartment. It is pertinent to inquire why this method of interpretation succeeds as well as it does[6] and to what extent information could be obtained, by more precise experiments, concerning the nature of the statistical distribution function describing the variable cellular properties. Because of the large number of the small compartments (3–5×10^9 per cc) the sums in Eqs. (12), (15), and (20) become integrals.

The transition is effected with the aid of a step function $A(\alpha)$ which has the following properties (Fig. 7):

$$\begin{aligned} A(\alpha) &= 0 && \text{for} \quad \alpha < \alpha_1, \\ &= A_1 && \text{for} \quad \alpha_1 \leq \alpha < \alpha_2, \\ &= A_1+A_2 && \text{for} \quad \alpha_2 \leq \alpha < \alpha_3, \qquad (25) \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ &= A_1+A_2+\cdots A_n && \text{for} \quad \alpha_n \leq \alpha. \end{aligned}$$

We may define $g(\lambda)$ by means of the Stieltjes integral[14]

[14] D. V. Widder, *The Laplace Transform* (Princeton University Press, Princeton, New Jersey, 1946).

of the function $A(\alpha)$,

$$g(\lambda)=1-\int_0^\infty (\lambda-\alpha)^{-1}dA(\alpha). \tag{26}$$

In like fashion we may define a step function $X(u)$ (Fig. 7):

$$\begin{aligned} X(u)&=0 && \text{for} \quad u<\lambda_1,\\ &=x_1 && \text{for} \quad \lambda_1\leq u<\lambda_2,\\ &=x_1+x_2 && \text{for} \quad \lambda_2\leq u<\lambda_3,\\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ &=x_1+x_2+\cdots x_n && \text{for} \quad \lambda_n\leq u. \end{aligned} \tag{27}$$

Hence, we have

$$[\lambda g(\lambda)]^{-1}=x_0/\lambda+\int_0^\infty (\lambda-u)^{-1}dX(u), \tag{28}$$

and

$$a_0(t)/a_0(0)=x_0+\int_0^\infty e^{-ut}dX(u). \tag{29}$$

The expressions are now in a form to permit generalization to the case $n\to\infty$.

If $X(u)$ were known then $[a_0(t)/a_0(0)]-x_0$ could be determined as its laplace transform. This would require the insertion into Eq. (26) of some known functional relation $A(\alpha)$ between the exchange rates. The solution of this equation would then be inserted into the integral Eq. (28) to determine the required $X(u)$. In practice neither the determination of $X(u)$ given $A(\alpha)$ nor the determination of $A(\alpha)$ given $X(u)$ is apt to be simple, and closed expressions in terms of elementary functions can be expected in very exceptional cases at best. However, considerable information can be obtained concerning the behavior of $[a_0(t)/a_0(0)]-x_0$ for small values of t. This requires a relation between the moments of $X(u)$ and $A(\alpha)$, which is derived in Appendix C.

Initial Behavior of $a_0(t)$

For small values of t Eq. (29) is equivalent to

$$\{[a_0(t)/a_0(0)]-x_0\}e^{xt}=\int_0^\infty e^{-(u-x)t}dX(u), \tag{30}$$

where x is the arithmetic mean of the X distribution about the origin (see Appendix). If we expand the exponential under the integral sign in powers of t and then integrate, the coefficients of the powers of t are the moments ν_r of the function $X(u)$ about the mean value x; thus,

$$[a_0(t)/a_0(0)-x_0]e^{xt}=X_0[1+(1/2)\nu_2t^2-(1/6)\nu_3t^3+\cdots],$$

and

$$[a_0(t)/a_0(0)-x_0]/X_0=e^{-xt}[1+(1/2)\nu_2t^2-(1/6)\nu_3t^3+\cdots],$$

so that in a semilog plot the initial slope is entirely determined by the mean x of the X distribution. Thus, as shown in the Appendix, the system behaves initially as though all the peripheral compartments were combined. If we write

$$\log\frac{a_0(t)}{a_0(0)}-x_0=\log X_0-xt+c(t), \tag{31}$$

$c(t)$ represents the deviation from linearity of the semilog plot of $[a_0(t)/a_0(0)]-x_0$. This deviation is initially zero but increases with time. To determine the deviation we take the series expansion

$$e^{c(t)}=1+(1/2)\nu_2t^2-(1/6)\nu_3t^3+\cdots,$$

and by successive differentiations

$$\begin{aligned} c(0)&=c'(0)=0,\\ c''(0)&=\nu_2,\\ c'''(0)&=-\nu_3,\\ c^{iv}(0)&=\nu_4-3\nu_2^2. \end{aligned} \tag{32}$$

Expanding $c(t)$ around the point $t=0$,

$$c(t)=(\nu_2t^2/2)-(\nu_3t^3/6)+[(\nu_4-3\nu_2^2)t^4/24]+\cdots, \tag{33}$$

and in terms of the moments of the B function derived in the Appendix,

$$c(t)=(b_2t^2/2x_0)-[b_3-xb_2(1-x_0)](t^3/6x_0)+\cdots. \tag{34}$$

To illustrate the magnitude of the deviation from true exponential behavior we consider the case where $B(\alpha)$ is a symmetrical ($b_3=0$) gaussian function whose standard deviation relative to the mean is 30 percent, i.e., $b_2=(0.3)^2b^2=0.09x_0^2x^2$. Taking for x_0 a typical value 0.3

$$\begin{aligned} c(t)&=\frac{b_2t^2}{2x_0}[1+xt(1-x_0)/3]\\ &=0.09\times0.15x^2t^2+(0.09\times0.7\times0.3/6)x^3t^3\\ &=0.0135x^2t^2+0.0031x^3t^3. \end{aligned}$$

For $tx<0.69$, i.e., up to the first half-value time, the deviation is less than 1 percent. This indicates the insensitivity of the result to a relatively large spread in the B function and thus indirectly to a spread in the properties of the peripheral compartments. Initially the deviation is controlled almost entirely by the second moment of the B function (taken about the mean) and is such that the curve is parabolic and convex to the axis of the abscissas. As time progresses the higher moments become effective. These are determined by the more detailed aspects of the distribution such as skew and excess. The convergence of $c(t)$ depends to a considerable extent on the magnitude of the higher moments. Even in the case of a gaussian distribution these may become quite large. Thus the approximation is usually only good for small t.

THE CATENARY SYSTEM

This system is of particular interest because of its importance in metabolic studies.[1,15,16] Here again (Fig. 3), although we have a system of n compartments, at most $2n-2$ of the dS_{jk}'s are to be determined, and thus only two species of S (radioactive and nonradioactive) are required. Introducing $\mathbf{R}_j$ for the total radioactivity in compartment j the equations are

$$\begin{aligned} a_{j-1}dS_{j\,j-1}+a_{j+1}dS_{j\,j+1}-a_j(dS_{j-1\,j}-dS_{j+1\,j})&=d\mathbf{R}_j \quad (35)\\ dS_{j\,j-1}+dS_{j\,j+1}-dS_{j-1\,j}-dS_{j+1\,j}&=dS_j, \end{aligned}$$

where

$$1\leq j\leq n \quad \text{and} \quad dS_{10}=dS_{01}=dS_{n\,n+1}=dS_{n+1\,n}=0.$$

The solution is

$$dS_{j\,j+1}=\frac{\begin{vmatrix}\sum_1^j d\mathbf{R}_k & a_j\\ \sum_1^j dS_k & 1\end{vmatrix}}{\begin{vmatrix}a_{j+1} & a_j\\ 1 & 1\end{vmatrix}}=\frac{\sum_1^j d\mathbf{R}_k-a_j\sum_1^j dS_k}{a_{j+1}-a_j}$$

$$dS_{j+1\,j}=\frac{\begin{vmatrix}a_{j+1} & \sum_1^j d\mathbf{R}_k\\ 1 & \sum_1^j dS_k\end{vmatrix}}{\begin{vmatrix}a_{j+1} & a_j\\ 1 & 1\end{vmatrix}}=\frac{a_{j+1}\sum_1^j dS_k-\sum_1^j d\mathbf{R}_k}{a_{j+1}-a_j}. \qquad (36)$$

Since $$\sum_1^n dS_k=\sum_1^n d\mathbf{R}_k=0,$$

$$dS_{j\,j+1}=\frac{a_j\sum_{j+1}^n dS_k-\sum_{j+1}^n d\mathbf{R}_k}{a_{j+1}-a_j}$$

$$dS_{j+1\,j}=\frac{\sum_{j+1}^n d\mathbf{R}_k-a_{j+1}\sum_{j+1}^n dS_k}{a_{j+1}-a_j}.$$

Thus the $dS_{j+1\,j}$, etc. are obtainable from determinations of the specific activities in the two adjoining compartments only, and of the total change in S and $\mathbf{R}$ for the entire system on one side, or the other of the boundary separating them.

[15] Zilversmit, Entenman, and Fishler, J. Gen. Physiol. **26**, 325 (1942).

[16] H. Branson, Science **106**, 404 (1947).

The Quasi-Steady State

Often the predicted behavior of a system with a given distribution of S_j and dS_{jk} is required. In a catenary system if only steady state problems are considered involving exchange between compartments then all studies of turnover in metabolic systems are excluded. On the other hand, if the most general case is considered where no pair of opposing transfers are equal, the solution becomes unnecessarily complex. We therefore consider an intermediate "quasi-steady state" situation which can be reduced either to the case of pure exchange or the case of uniform turnover. Here it is assumed that the contents of compartments 1 and n (Fig. 3) are very large so that S_1 and S_n are nearly constant. S is exchanging at equal and opposite rates ρ_j between compartments j and $j+1$, and superimposed on this is an over-all "turnover rate" τ. Thus

$$dS_{j\,j+1}/dt=\rho_j \quad \text{and} \quad dS_{j+1\,j}/dt=\rho_j+\tau.$$

The differential equations for this system are

$$\begin{aligned} da_1/dt&=1/S_1[a_2\rho_1-a_1(\rho_1+\tau)]\\ &\cdots\cdots\cdots\cdots\\ da_k/dt&=1/S_k[a_{k-1}(\rho_{k-1}+\tau)\\ &\qquad +a_{k+1}\rho_k-a_k(\rho_{k-1}+\rho_k+\tau)] \qquad (37)\\ &\cdots\cdots\cdots\cdots\\ da_n/dt&=1/S_n[a_{n-1}(\rho_{n-1}+\tau)-a_n\rho_{n-1}]. \end{aligned}$$

Expressing the solution as a sum of exponentials we find the λ's are the roots of the determinant relation.

$$D_n(\lambda)=\begin{vmatrix} d_1 & e_1 & 0 & 0 & \cdot & \cdot & \cdot & \cdot & 0\\ f_1 & d_2 & e_2 & 0 & \cdot & \cdot & \cdot & \cdot & 0\\ 0 & f_2 & d_3 & e_3 & \cdot & \cdot & \cdot & \cdot & 0\\ 0 & 0 & f_3 & d_4 & e_4 & \cdot & \cdot & \cdot & 0\\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0\\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0\\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & f_{n-2} & d_{n-1} & e_{n-1}\\ 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 & f_{n-1} & d_n \end{vmatrix}=0 \qquad (38)$$

where

$$\begin{aligned} d_j&=[\lambda S_j-(\rho_{j-1}+\rho_j+\tau)], \quad e_j=\rho_j\\ f_j&=\rho_j+\tau, \quad \rho_0=(\rho_n+\tau)=0. \end{aligned}$$

We define D_j as the determinant of rank j having the same elements as in the first j rows and columns of D_n. By successive expansions on the elements of the right-hand column,

$$D_j=d_jD_{j-1}-e_jf_jD_{j-2}. \qquad (39)$$

The first special case occurs if the ρ's are zero. Since now $e_j=0$ the relation becomes

$$D_n=d_1d_2\cdots d_n=0, \qquad (40)$$

with roots

$$\lambda_i=\tau/S_i.$$

Under these conditions the Bateman equations for radioactive decay apply. The x coefficients (see mammillary case) depend upon the various possible types of boundary conditions, and are derived for four different

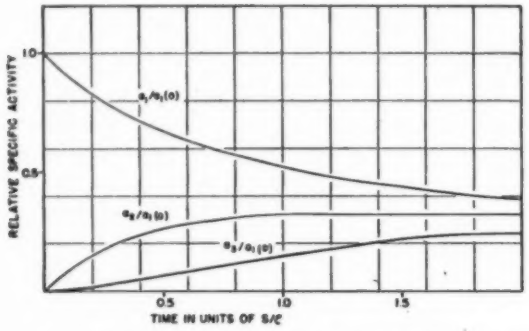


FIG. 8. Relative specific activity changes in a typical three-compartment catenary system.

sets of initial conditions in the classical treatise on radioactivity by Rutherford, *et al.*[17] Strictly speaking, the equations are in the Bateman form only if the dependent variables are the $\mathbf{R}_j$ rather than the a_j, but the result differs only in the factor S_j.

The case where $\tau=0$ and $e_j=f_j$ represents mixing by processes of exchange between compartments. In the case of more than three compartments the λ's will depend upon the order in which the exchange rates occur in the chain. The four-compartment case will serve as an illustration, thus

$$D_3=d_1d_2d_3d_4[1-(e_2^2/d_2d_3)-(e_1^2/d_1d_2)-(e_3^2/d_3d_4)-(e_1^2e_3^2/d_1d_2d_3d_4)]=0. \quad (41)$$

This is not a symmetrical function since it lacks the terms in $e_1^2e_2^2$ and $e_2^2e_3^2$. Thus different solutions will be obtained depending upon the permutations of the numerical subscripts. Here in contrast to the mammillary case the solution depends upon the order in which the compartments occur.

Exchange Processes in a Three-Compartment Catenary System

The roots are the same in this case as in the three-compartment mammillary system. However, the x coefficients will differ since the important cases will generally involve different initial conditions. As an illustration we will take the case where the exchange rates and contents of the compartments approach one another. Here

$$\lambda_1\rightarrow\tfrac{1}{2}(\rho_1/S_1+\rho_2/S_3)$$
$$\lambda_2\rightarrow\lambda_1+(\rho_1+\rho_2)/S_2.$$

The solution is the same as that for the three-compartment mammillary system except that the compartments are numbered differently. For the case where all of the radioactivity is initially in an end compartment (compartment 1) a close approximation is

$$\begin{aligned} a_1/a_1(0)&=S_1/S+\tfrac{1}{2}e^{-\lambda_1 t}+(\tfrac{1}{2}-S_1/S)e^{-\lambda_2 t}\\ a_2/a_1(0)&=S_1/S(1-e^{-\lambda_2 t})\\ a_3/a_1(0)&=S_1/S-\tfrac{1}{2}e^{-\lambda_1 t}+(\tfrac{1}{2}-S_1/S)e^{-\lambda_2 t}. \end{aligned} \quad (42)$$

[17] Rutherford, Chadwick, and Ellis, *Radiations from Radioactive Substances* (The Macmillan Company, New York, 1930).

Figure 8 shows graphically the time variation of the specific activities. The initial variation of a_1 resembles that of the central compartment of a three-compartment mammillary system possessing widely different exchange rates and might easily be confused with this if observations were confined to compartment 1 alone.

CONCLUSION

There is a definite limit to the amount of information which can be obtained from a completely general treatment of isotope problems. Thus, although there is a basic interfusion equation, its practical use in systems of uniformly variable composition must await the development of satisfactory methods for determining the uniformly variable radioactivity and concentration of S in such systems. We are thus already limited to the less general case of systems of compartments.

Here the advantage of maximum generality has been demonstrated in the general conclusions which can be drawn as to the necessary number of species of S required and the minimum observations which must be made. Unfortunately the conclusions emphasize the fundamental difficulties encountered in the use of isotopes in physiology because of the practical limitations on the isolation of compartments and establishment of their uniformity. An interesting conclusion is that the study of exchange between erythrocytes and plasma *in vivo* can be correctly achieved with a single radioactive isotope. However, one must obtain a thorough record of the specific activities of cells and plasma and of the changes of radioactivity in the cells during the experimental period. Such an experiment is possible in the complex mammalian case because the cells and plasma form two compartments at one end of a catenary system. As we have indicated, the lumping of the variable population of cells into a single effective compartment is a satisfactory approximation.

The study of the disappearance of labeled material from the plasma is considerably more difficult to interpret on a rigorous basis. First, if the system is a mammillary one, observations in the plasma alone are not sufficient to determine whether there are several peripheral compartments or only two or three because of the insensitivity of the results to the lumping of many compartments into a few average ones. If the plasma were the terminal compartment of a catenary system, its behavior might well be similar to that observed in the mammillary case. In all probability neither simple picture will apply.

In considering the problem of exchange in the catenary system the analysis becomes considerably more complicated because of the dependence of the exponential constants on the order of the compartments. As a result the problems will be more satisfactorily treated on an individual basis perhaps with approximation methods or by the use of electrical analogies. Fortunately the turnover problems are of more immediate interest and these are readily handled in the general

case using the Bateman equations. Although open systems are not discussed they can usually be included by the addition of one or more extra compartments. Interesting cases arise where the quasi-steady state treatment does not suffice, and the amounts of S in the compartments vary. Typical of these are the idealized interpretations of the excretion of Na^{22} by Burch, Threefoot, and Cronvich.[18] Some of these problems will doubtless be treated by Branson's method.[16] The discussion of other cases must be postponed for the present.

ACKNOWLEDGMENTS

An earlier version of this paper was read and commented upon by Dr. Cornelius Tobias and Dr. L. J. Savage. We wish to thank Dr. Savage sincerely for the extreme thoroughness of his review. Dr. W. E. Cohn also provided several constructive comments.

APPENDIX A

Derivation of the Interfusion Equation

Consider a differential element of area $dxdy$ at the origin taken perpendicularly to the gradient of A so that A varies only with z (Fig. 9). At a nearby point P locate a volume element dv whose coordinates are x, y, z. Of the Ndv particles emitted by this element per unit time the fraction $Fdxdy$ will cross the element of area. If dv is sufficiently close to the origin the impinging particles will not collide with other molecules of S and thus lose their identity. The fraction labeled will thus be

$$A(z)+(\partial A/\partial z)z$$

and the number of labeled particles crossing the element of area in dt seconds is

$$N(x, y, z)F(x, y, z)[A(z)+(\partial A/\partial z)z]dvdxdydt.$$

Integrating over x, y, and z for the entire half space we obtain for the total particles traversing the area after their last interaction

$$dxdydt\int\int\int NF[A(z)+(\partial A/\partial z)z]dv.$$

Integration over the other half space yields the same expression except that $(\partial A/\partial z)z$ has a negative sign, and the net transport

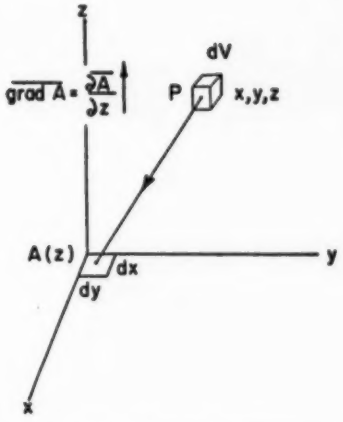


FIG. 9. Derivation of the interfusion equation.

of labeled particles per area unit is obtained by subtraction;

$$\mathbf{dN^*/dt=\partial A/\partial z}\int\int\int \mathbf{NFzdv=\rho z\ grad}\ A,$$

where $\rho=\int\int\int NFdv$, i.e., the total exchange rate across the surface element. The quantity $\mathbf{z}$ represents the mean distance from which particles originate in traversing the area. In the special case of self-diffusion of labeled gas molecules in a box under uniform pressure and temperature it is 4/3 times the mean free path. However, in the general case it may depend upon the nature of transport processes by which variable concentration gradients may be maintained in the system. Since the partial derivative is removed from under the integral sign, it must not vary appreciably over distances of the order of $\mathbf{z}$. This is similar to the analogous case in classical kinetic theory where the macroscopic treatment breaks down as soon as dimensions of the order of the mean free path become important.

Both ρ and $\mathbf{z}$ vary only with the space coordinates and not with time, thus their product will be constant at a given point in space and can be called the interfusion constant, I. Although, in the more familiar diffusion theory, D is usually constant it is not necessarily so and both D and I in the general case may be scalar point functions.

So far, attention has been confined to only one type of S. If several forms of chemical combination of the substance are to be considered then, in addition to the location of a labeled particle in space by the coordinates x, y, and z, we must also add one more coordinate by which the state of chemical combination is to be described.

APPENDIX B

The Three-Compartment Mammillary System

To illustrate the application of Eqs. (13) and (22) to a special case we will consider a system where $n=2$. Equation (13) becomes

$$g(\lambda)=1-A_1/(\lambda-\alpha_1)-A_2/(\lambda-\alpha_2)=0$$

or

$$\lambda^2-\lambda(\alpha_1+\alpha_2+A_1+A_2)+A_1\alpha_2+A_2\alpha_1+\alpha_1\alpha_2=0.$$

Thus,

$$\lambda=\frac{(\alpha_1+\alpha_2+A_1+A_2)\pm[(\alpha_1+\alpha_2+A_1+A_2)^2-4(A_1\alpha_2+A_2\alpha_1+\alpha_1\alpha_2)]^{\frac{1}{2}}}{2}$$

$$=\frac{(\alpha_1+\alpha_2+A_1+A_2)\pm[(\alpha_1-\alpha_2)^2+(A_1+A_2)^2+2(\alpha_1-\alpha_2)(A_1-A_2)]^{\frac{1}{2}}}{2}$$

$$=\alpha+(\Sigma A/2)\pm\tfrac{1}{2}[\Delta\alpha^2+(\Sigma A)^2-2\Delta\alpha\Delta A]^{\frac{1}{2}}$$

$$=\alpha+(\Sigma A/2)\pm\tfrac{1}{2}D^{\frac{1}{2}},$$

[18] Burch, Threefoot, and Cronvich, J. Lab. and Clin. Med. 34, 14 (1949).

α being the mean of the exchange rates relative to the contents of the peripheral compartments and ΣA the total exchange rate relative to the central one. D is the discriminant of the quadratic equation.

From Eq. (22) (inserting $\lambda_0=0$ for completeness),

$$x_1=\frac{(\lambda_1-\alpha_1)(\lambda_1-\alpha_2)}{(\lambda_2-\lambda_1)(\lambda_0-\lambda_1)}\,x_2=\frac{(\lambda_2-\alpha_1)(\lambda_2-\alpha_2)}{(\lambda_1-\lambda_2)(\lambda_0-\lambda_2)}.$$

Thus

$$x_1=-\frac{[-\Delta\alpha+\Sigma A-D^{\frac{1}{2}}][\Delta\alpha+\Sigma A-D^{\frac{1}{2}}]}{4[\alpha+(\Sigma A/2)-(D^{\frac{1}{2}}/2)]D^{\frac{1}{2}}}.$$

$$x_2=\frac{[-\Delta\alpha+\Sigma A+D^{\frac{1}{2}}][\Delta\alpha+\Sigma A+D^{\frac{1}{2}}]}{4[\alpha+(\Sigma A/2)+(D^{\frac{1}{2}}/2)]D^{\frac{1}{2}}}$$

as $\alpha_1\to\alpha_2$

$$x_1\to\Delta\alpha^2/4\alpha\Sigma A$$

$$x_2\to\Sigma A/(\alpha+\Sigma A).$$

APPENDIX C

The Moments of the Distribution Functions

Since $X(u)$ is dimensionless and $A(\alpha)$ is not, it is more convenient to consider the related function $B(\alpha)$ defined by

$$dA(\alpha)=\alpha dB(\alpha), \tag{i}$$

with

$$B(0)=0.$$

In place of Eq. (26) we write

$$g(\lambda)=1-\int_0^\infty \alpha(\lambda-\alpha)^{-1}dB(\alpha). \tag{ii}$$

We define the moments B_r of the function $B(\alpha)$ about the origin by

$$B_r=\int_0^\infty \alpha^r dB(\alpha)\quad(r=0,1,2,\cdots). \tag{iii}$$

Then $g(\lambda)$ can be expanded formally in powers of λ^{-1}:

$$g(\lambda)=1-B_1/\lambda-B_2/\lambda^2\cdots. \tag{iv}$$

We note also from Eqs. (ii) and (iii)

$$g(0)=1+B_0=x_0^{-1}. \tag{v}$$

In like manner, if we define the moments about the origin,

$$X_r=\int_0^\infty u^r dX(u), \tag{vi}$$

expanding the integral,

$$\int_0^\infty(\lambda-u)^{-1}dX(u)=\lambda^{-1}[X_0+X_1/\lambda+X_2/\lambda^2+\cdots]. \tag{vii}$$

Hence, Eq. (28) can be written formally

$$\begin{aligned}1/g(\lambda)&=x_0+X_0+X_1/\lambda+X_2/\lambda^2+\cdots\\&=1+X_1/\lambda+X_2/\lambda^2+\cdots,\end{aligned} \tag{viii}$$

since

$$x_0+X_0=1.$$

If we multiply together Eqs. (iv) and (viii) we obtain after a minor reduction

$$0=(X_1-B_1)\lambda+(X_2-B_1X_1-B_2)/\lambda^2+\cdots,$$

and since the right member must vanish identically in λ, we obtain the recursion relations

$$\begin{aligned}&X_1=B_1,\\&X_2=B_1X_1+B_2,\\&\quad\cdots\cdots\cdots\\&X_r=B_1X_{r-1}+B_2X_{r-2}+\cdots+B_r.\end{aligned} \tag{ix}$$

The moments about the mean are usually more convenient. Let x and b represent the means of the two distributions X and B. Then

$$X_1=X_0x,\quad B_1=B_0b.$$

Likewise let ν_r and b_r designate the moments about the mean. These are related to the moments X_r and B_r, respectively, by the relations

$$X_r=X_0\left[\nu_r+\binom{r}{1}x\nu_{r-1}+\cdots+\binom{r}{2}x^{r-2}\nu_2+x^r\right] \tag{x}$$

and

$$B_r=B_0\left[b_r+\binom{r}{1}bb_{r-1}+\cdots+\binom{r}{2}b^{r-2}b_2+b^r\right] \tag{xi}$$

since $\nu_1=b_1=0$. These can be substituted into (ix) obtaining a series of relations between the means and the moments about the mean. Recalling that

$$X_0=1-x_0=B_0/(1+B_0),$$

we write down only the following:

$$\begin{aligned}b&=x_0x,\\b_2&=x_0\nu_2,\\b_3&=x_0\nu_3+B_0bb_2.\end{aligned} \tag{xii}$$

x can be expressed in terms of the system parameters, thus

$$x=b/x_0=B_1/B_0x_0=S\Sigma\rho/S_0(S-S_0)=\Sigma\rho[1/S_0+1/(S-S_0)].$$

Comparing with Eqs. (8), x becomes entirely equivalent to the exponential coefficient of a two-compartment system in which the exchange rate is the sum of the rates for the peripheral compartments. One equivalent compartment contains the sum of the contents of the peripheral compartments.

Erratum: Diffraction Pattern in a Circular Aperture Measured in the Microwave Region

[J. Appl. Phys. 21, 761–767 (1950)]

C. L. ANDREWS
General Electric Research Laboratory, The Knolls, Schenectady, New York

DR. Hans Severin has called the author's attention to an error made in describing Severin's measurements of diffraction by a circular aperture in a metal screen. The author said that Severin coated the front and back of his metal screen with a "paint" that reduced reflections to 10 percent of metallic reflection. Actually Severin's apertures were cut in perfectly reflecting metal screens of 1 meter² or more. The coated screens flanked the metal diffracting screens at an oblique angle of about 160 degrees.

Electric and Magnetic Forces between Sphere and Wire

W. R. SMYTHE
California Institute of Technology, Pasadena, California
(Received December 18, 1950)

Formulas are derived for the force and scalar potential of a dielectric sphere in the field of a charged wire and for the force and vector potential of a permeable sphere and a circular cylinder carrying current. When the sphere is conducting the electric force is $q^2a^2(\pi\epsilon_v b)^{-1}(b^2-a^2)^{-\frac{1}{2}}\sin^{-1}(a/b)$ in mks units, where q is the charge per unit length on the wire, a the sphere radius and b the distance from its center to the wire. The same formula, with i for q and μ_v^{-1} for ϵ_v applies to a sphere of large permeability in the magnetic field of a circular cylinder carrying a current i.

OCCASIONALLY one needs to know the force between an uncharged dielectric or conducting sphere and a charged wire or between a permeable sphere and a wire carrying current. These forces probably account in part for the accumulation of oil and magnetic dust on electric wiring in stagnant air. Although simple formulas in closed form exist in both cases when the capacitivity or permeability of the sphere is large we have been unable to find them in the literature. In the electric case the sphere may be any size but the wire must be thin enough so that the charge on it may be taken uniform. In the magnetic case there is no such restriction and the results are rigorous for any size sphere and any circular cylinder or tube in which the current is axially symmetrical, provided sphere and tube do not intersect.

For both types of field the coordinate system of Fig. 1 is used. The origin is taken at the center of the sphere and both the cylindrical coordinate system ρ, ϕ, z and the spherical polar system r, θ, ϕ are used. The field point is P. The rationalized mks system of units is used throughout, K being written for ϵ/ϵ_v and K_m for μ/μ_v.

POTENTIALS IN ELECTROSTATIC CASE

The potential of a line at $\rho=b$, parallel to the z-axis and uniformly electrified with a charge q per unit length is, if the zero of potential is taken at $\rho=0$,

$$V=\frac{-q}{4\pi\epsilon_v}\ln\left[1+\left(\frac{\rho}{b}\right)^2-\frac{2\rho}{b}\cos\phi\right]. \tag{1}$$

When $\rho<b$ this may be written by Dwight[1] 418

$$V=\frac{q}{2\pi\epsilon_v}\sum_{n=1}^{\infty}\frac{1}{n}\left(\frac{\rho}{b}\right)^n\cos n\phi. \tag{2}$$

In spherical polar coordinates this is

$$V=\frac{q}{2\pi\epsilon_v}\sum_{n=1}^{\infty}\left(\frac{r}{b}\right)^n\frac{\sin^n\theta}{n}\cos n\phi. \tag{3}$$

The potential inside the sphere must be everywhere finite and, to match (3), must depend on θ and ϕ in the same way so that

$$V_i=\frac{q}{2\pi\epsilon_v}\sum_{n=1}^{\infty}A_n\left(\frac{r}{b}\right)^n\frac{\sin^n\theta}{n}\cos n\phi. \tag{4}$$

The potential V_0 outside must reduce to (4) at $r=a$ and must contain, in addition to (3), a term due to the polarization that vanishes at $r=\infty$. If $r^nS_n(\theta,\phi)$ is a solution of Laplace's equation then $r^{-n-1}S_n(\theta,\phi)$ is also a solution[2] so that

$$V_0=\frac{q}{2\pi\epsilon_v}\sum_{n=1}^{\infty}\left[\left(\frac{r}{b}\right)^n+\frac{(A_n-1)a^{2n+1}}{b^nr^{n+1}}\right]\frac{\sin^n\theta}{n}\cos n\phi. \tag{5}$$

The condition that the normal displacement, $\epsilon\partial V/\partial r$, be continuous across the boundary at $r=a$ gives

$$A_n=\frac{2n+1}{Kn+n+1}. \tag{6}$$

For a conducting sphere K is infinite so that A_n is zero and the polarization term V_c' in V_0 becomes

$$V_c'=\frac{-qa}{2\pi\epsilon_v r}\sum_{n=1}^{\infty}\frac{1}{n}\left(\frac{a^2\sin\theta}{br}\right)^n\cos n\phi. \tag{7}$$

The summation factor is identical with that in (2) with $a^2\sin\theta/(br)$ in place of ρ/b so that, comparison with (1) gives for the potential of the induced charge on a conducting sphere

$$V_c'=\frac{qa}{4\pi\epsilon_v r}\ln\left[1+\frac{a^4\sin^2\theta}{b^2r^2}-\frac{2a^2\sin\theta}{br}\cos\phi\right]. \tag{8}$$

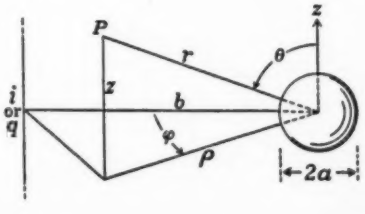


FIG. 1.

[1] H. B. Dwight, *Tables of Integrals and Other Mathematical Data* (Macmillan Company, New York, 1947), revised edition. The serial number of the relevant formula is given.

[2] W. R. Smythe, *Static and Dynamic Electricity* (McGraw-Hill Book Company, Inc., New York, 1950), Second edition, p. 130.

This, added to (1), gives the total potential outside the conducting sphere. As r increases, V_c' approaches the potential of a dipole of moment $2qa^2/b$ placed at the center of the sphere and pointing away from the wire. There is no q/r term in V_c' so the sphere is uncharged.

FORCE IN ELECTROSTATIC CASE

The force on the sphere is most easily found from its reaction on the wire by multiplication of the ρ-component of its polarization field at $\rho=b$, $\phi=0$ by qdz and integration from $-\infty$ to $+\infty$. Writing the second term of (5) in terms of ρ and z at $\phi=0$, this gives

$$F=\frac{-q^2}{2\pi\epsilon_v}\sum_{n=1}^{\infty}\frac{(K-1)a^{2n+1}}{(Kn+n+1)b^n}\frac{\partial}{\partial\rho}\int_{-\infty}^{\infty}\frac{\rho^n dz}{(\rho^2+z^2)^{n+\frac{1}{2}}}. \quad (9)$$

This is integrated by setting $x=z^2/\rho^2$ in Dwight 855.2. Substitution of b for ρ in the result gives

$$F=\frac{q^2}{\pi\epsilon_v}\sum_{n=1}^{\infty}\frac{2\cdot4\cdots(2n-2)n(K-1)}{1\cdot3\cdots(2n-1)(Kn+n+1)}\left(\frac{a}{b}\right)^{2n+1}. \quad (10)$$

For the conducting sphere $K=\infty$ and the sum of the series is $a^2b^{-1}(b^2-a^2)^{-\frac{1}{2}}\tan^{-1}[a(b^2-a^2)^{-\frac{1}{2}}]$ by Dwight 505.4. Writing in terms of the arcsine gives

$$F_c=\frac{q^2a^2}{\pi\epsilon_v b(b^2-a^2)^{\frac{1}{2}}}\sin^{-1}\frac{a}{b} \quad (11)$$

where q is the charge per *unit length* on the wire.

POTENTIALS IN MAGNETIC CASE

The magnetic vector potential A_z of the wire in Fig. 1 is identical with (1), (2), and (3) except that i and μ_v^{-1} replace q and ϵ_v. Upon this must be superimposed a vector potential due to the magnetization which has zero divergence and therefore contains a vector surface harmonic.[3] Its θ-component, to combine with that of A_z should contain $r^n\cos n\phi$ where $r<a$ and so the legendre function is $P_n^n(\cos\theta)$ or $(2n-1)!!\sin^n\theta$.[4] Setting $(2n-1)!!C=n$, $D=0$, $n=m$ and $\delta=\frac{1}{2}\pi$ in the spherical vector solution[5] we obtain for the surface vector harmonic $\mathbf{S}_n(\theta,\phi)$, using $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ for unit vectors,

$$\boldsymbol{\theta}\sin^{n-1}\theta\cos n\phi-\boldsymbol{\phi}\sin^{n-1}\theta\cos\theta\sin n\phi. \quad (12)$$

This gives for the magnetization potentials $\mathbf{A}'$

$$\mathbf{A}_i'=\frac{\mu_v i}{2\pi}\sum_{n=1}^{\infty}C_n\left(\frac{r}{b}\right)^n\mathbf{S}_n(\theta,\phi)\quad r<a, \quad (13)$$

$$\mathbf{A}_0'=\frac{\mu_v i}{2\pi}\sum_{n=1}^{\infty}C_n\frac{a^{2n+1}}{b^n r^{n+1}}\mathbf{S}_n(\theta,\phi)\quad r>a. \quad (14)$$

[3] See reference 2, p. 151, Eq. (7) and p. 267, Eq. (2)
[4] See reference 2, p. 149, Eq. (14).
[5] See reference 2, p. 267, Eq. (2).

Clearly $\mathbf{A}_i'$ is finite inside the sphere, $\mathbf{A}_0'$ vanishes at $r=\infty$, and $\mathbf{A}_i'$ equals $\mathbf{A}_0'$ at its surface. For the total potential we add A_z to both (13) and (14). It remains, by a suitable choice of C_n, to match H_ϕ and H_θ at the boundary $r=a$. The magnetic flux densities are given by

$$B_\phi=\frac{\partial(rA_\theta')}{r\partial r}-\frac{\partial A_z}{\partial\rho},\quad rB_\theta=\cot\theta\frac{\partial A_z}{\partial\phi}-\frac{\partial(rA_\phi')}{\partial r}. \quad (15)$$

Both the H_θ match and the H_ϕ match give

$$C_n=\frac{1-K_m}{K_m n+n+1}. \quad (16)$$

When $K_m=\mu/\mu_v$ is very large, C_n becomes $-1/n$ so

$$A_\theta'=\frac{-\mu_v ia}{4\pi r\sin\theta}\sum_{n=1}^{\infty}\frac{1}{n}\left(\frac{a^2\sin\theta}{br}\right)^n\cos n\phi,$$

$$A_\phi'=\frac{\mu_v ia\cot\theta}{2\pi}\sum_{n=1}^{\infty}\frac{1}{n}\left(\frac{a^2\sin\theta}{br}\right)^n\sin n\phi,$$

outside the sphere. The first of these can be summed as (7) was summed and the second by Dwight 421. Thus outside a very permeable sphere its magnetization potential is

$$A_\theta'=\frac{\mu_v ia}{4\pi r\sin\theta}\ln\left[1+\frac{a^4\sin^2\theta}{b^2r^2}-\frac{2a^2\sin\theta}{br}\cos\phi\right], \quad (17)$$

$$A_\phi'=\frac{\mu_v ia\cot\theta}{2\pi r}\tan^{-1}\frac{a^2\sin\theta\sin\phi}{br-a^2\sin\theta\cos\phi}. \quad (18)$$

FORCE IN THE MAGNETIC CASE

The attractive force between cylinder and sphere is the product of the transverse magnetic induction B_ϕ' of the sphere at the cylinder axis by idz integrated from $z=-\infty$ to $z=+\infty$. Thus

$$F=i\int_{-\infty}^{+\infty}B_\phi' dz=i\int_{-\infty}^{+\infty}\frac{\partial(rA_\theta')}{r\partial r}dz.$$

Substitution for A_θ' from (13) with $\phi=0$ gives

$$F=\frac{\mu_v i^2}{2\pi b}\sum_{n=1}^{\infty}\frac{(K_m-1)na^{2n+1}}{K_m n+n+1}\int_{-\infty}^{+\infty}\frac{dz}{(b^2+z^2)^{n+\frac{1}{2}}}.$$

Evaluation of this integral as with (9) gives

$$F=\frac{\mu_v i^2}{\pi}\sum_{n=1}^{\infty}\frac{2\cdot4\cdots(2n-2)n(K_m-1)}{1\cdot3\cdots(2n-1)(nK_m+n+1)}\left(\frac{a}{b}\right)^{2n+1}. \quad (19)$$

As with (9) this becomes, as $K_m=\mu/\mu_v$ approaches ∞

$$F_\infty=\frac{\mu_v i^2a^2}{\pi b(b^2-a^2)^{\frac{1}{2}}}\sin^{-1}\frac{a}{b}. \quad (20)$$

Letters to the Editor

Transmission Line Kerr Cell*

J. W. Beams and H. S. Morton, Jr.
University of Virginia, Charlottesville, Virginia
(Received January 21, 1951)

If plane-polarized light with the plane of polarization at an angle (45°) to the electrical field passes through a liquid showing the Kerr electro-optical effect, the component parallel to the field travels at a different velocity from the component perpendicular to the field so that the light emerges elliptically polarized. The phase retardation $D=2\pi BlE^2$, where l is the length of the light path through the electrical field E, and B is the Kerr constant which varies with different substances, temperatures, and wavelengths of the light used. In many liquids no time lag has been observed between the change in the field and the change in the Kerr electro-optical effect, so that this effect has been used widely to make a light shutter (Kerr cell) which responds almost instantaneously to electrical control.[1] Usually a Kerr cell consists of a parallel plate condenser immersed in a liquid and placed between crossed Nicol prisms or polaroids. If the field is on, light will pass the combination, but if the field is off, no light can pass through. For very fast acting Kerr cell light shutters, it has been necessary to select a liquid with a large Kerr constant B in order to get high light transmission (large D) at reasonable fields E. Also the dielectric constant must be small in order to reduce the circuital capacity to a minimum so that the cell may be operated with sufficient speed. In most applications, it has been necessary to use nitrobenzine in Kerr cells, as no other available transparent liquid possesses a large enough Kerr constant for practical use. Unfortunately, nitrobenzine is unstable chemically and is relatively conducting and relatively opaque in the violet. Furthermore, the dielectric constant is so high that the plates of the Kerr cell must be very small (which, in turn, gives rise to an inhomogeneous field) in order to avoid large circuital capacity.

The purpose of this report is to describe a very fast acting Kerr cell light shutter arrangement in which almost any liquid showing the Kerr effect may be used and in which the Kerr constant and dielectric constant are of much less importance than in the standard Kerr cell. In Fig. 1, AC is a parallel plate transmission

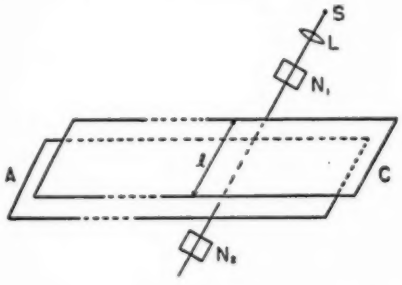


Fig. 1.

line with a transparent liquid showing the Kerr effect between the plates. The phenomena to be observed is at S, L is a lens, and N_1 and N_2 are Nicol prisms so oriented that the plane of polarization makes an angle of 45° with the lines of force between the plates. This arrangement clearly acts as a Kerr cell light shutter which can be opened and closed by electrical impulses on the transmission line AC. It will be observed that the velocity of an electrical impulse on such a line depends upon $(LC)^{-\frac{1}{2}}$, where L and C are the inductance and capacity per unit length. This ratio is independent of the width of the transmission line (neglecting edge corrections), so that the effective light path through the Kerr cell may be made as long as desired without introducing harmful capacity effects, etc. As a result, clear liquids which have high dielectric strength and are transparent in the ultraviolet or infrared as well as in the visible may be used instead of nitrobenzine. In our experiments the aluminum parallel plates of the transmission line were 335 cm long, 15.3 cm wide, and spaced 4.6 mm apart. The liquid between the plates was Halowax oil No. 1007, which has a dielectric constant of 5.13 at one megacycle. The total capacity of the line measured at 1000 cycles was 5.2×10^{-9} farads. The velocity of the electrical impulse on the line was 1.32×10^{10} cm/sec, and the surge impedance was about 5 ohms. Several pairs of strain-free glass windows were provided along AC, one pair being at C. So far the above arrangement has been used to study rapidly occurring phenomena at S. The Kerr cell was actuated with transients put on the line AC by a large capacity, low impedance, double spark gap impulse generator timed to open and close the cell at the desired instant. However, it is clear that the Kerr cell may be actuated by standing waves on the transmission line or that the shutter may be composed of two similar transmission lines of different length with their fields perpendicular, discharged at the same time, etc. The above arrangement will be described in more detail elsewhere.

* This work was supported by the Navy Bureau of Ordnance, Contract NOrd-7873.

[1] J. W. Beams, Revs. Modern Phys. 4, 135 (1932). For references see Rev. Sci. Instr. 1, 780 (1930).

Temporary Enhancement of Hysteresis in Barium Titanate Samples

D. R. Young
Electron Tube Research Laboratory, International Business Machines Corporation, Poughkeepsie, New York
(Received January 31, 1951)

Rzhanov[1] has recently reported aftereffects as a result of application of high ac fields to polycrystalline barium titanate samples at room temperatures. We have also found such effects and have observed, in addition, that they become more pronounced if the field is applied while the sample is heated above the Curie temperature and then cooled to room temperature (see Fig. 1). It can be seen that the maximum polarization is

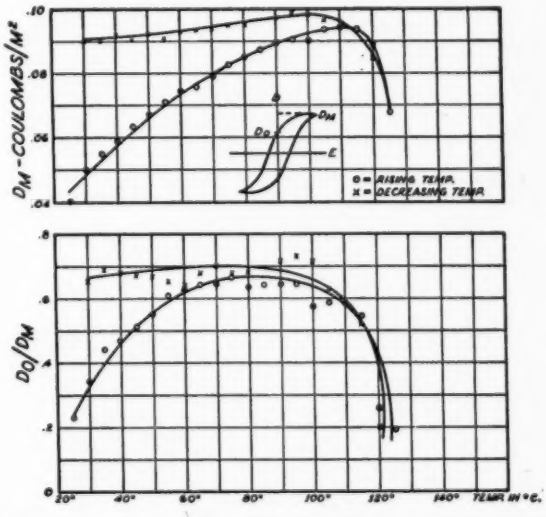


Fig. 1. D_M and D_O/D_M plotted as a function of temperature showing effects of treatment. Electric field strength $=8.34\times10^5$ volts per meter peak.

increased after treatment and that the hysteresis curve becomes more nearly rectangular.

If this is done with a dc field, the hysteresis curve appears highly unsymmetrical owing to the polarized nature of the dielectric; however, heating to only 50°C in the presence of an ac field

will restore symmetry, and the hysteresis characteristic will be the same as if it were treated by an ac field to 120°C. If the sample, after treatment, is heated to 400°C, without field applied, for one hour, it does not revert to its original state.

This modification is not permanent as can be seen in Fig. 2, where the decay is shown to occur in some hundreds of hours and

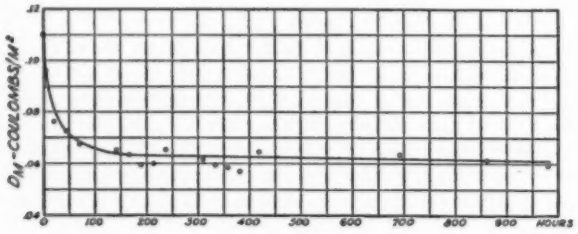


FIG. 2. Maximum D as a function of time. Field applied $=8.34\times10^6$ volts per meter.

varies from one sample to another. No dependence of decay rate on temperature has been found.

The author is indebted to Dr. Hans Jaffe of The Brush Development Company for calling his attention to the work of Rzhanov.

[1] A. V. Rzhanov, Zhur. Eksp. Teoret. Fiz. 19, 335–45 (1949).

Difference between Turbulence in a Two-Dimensional Fluid and in a Three-Dimensional Fluid

T. D. LEE
Department of Physics, University of California, Berkeley, California
(Received January 19, 1951)

THE difference between a two-dimensional and a three-dimensional fluid can easily be seen from the vorticity equation given as

$$\dot{\omega}+(\mathbf{v},\nabla)\omega=\nu\Delta\omega+(\omega,\nabla)\mathbf{v}, \tag{1}$$

where ω, $\mathbf{v}$, and ν are the vorticity, velocity, and kinematic viscosity of the fluid. In the two-dimensional case, $(\omega,\nabla)\mathbf{v}$ is identically zero. Hence, if one neglects viscosity in a system moving with the fluid, the vorticity never changes, and the scattering of energy between eddies does not lead to any change in vorticity. This conservation law forbids the fulfillment of an ergodic hypothesis for a two-dimensional fluid. Indeed, if one raises the same arguments to a two-dimensional case, as in the current theory of statistical turbulence,[1] a contradiction can readily be found.

On multiplying Eq. (1) by ω and transforming it to the wave number space, we can write

$$-\frac{\partial}{\partial t}\int_0^\infty F(k)k^2dk=2\nu\int_0^\infty F(k)k^4dk, \tag{2}$$

where $F(k)$ is the energy spectrum of a two-dimensional turbulence with k as the wave number. For definiteness, we shall use Heisenberg's[2] form of energy transfer between eddies which, for a steady state of turbulence, yields the form of $K(k)$ as

$$F(k)=\tfrac{1}{3}v_0^2k_0^{3/2}k^{-5/2} \quad \text{for} \quad k_0\leqslant k<k_s, \tag{3}$$

and

$$k_s/k_0=0.22(R_0\kappa)^{3/4},$$

where κ is the coefficient of eddy viscosity, and R_0 is the Reynolds' number given as $\pi v_0/\nu k_0$.

For a steady state of turbulence, we have

$$-\frac{\partial}{\partial t}\int_0^\infty F(k)k^2dk=-\frac{\partial}{\partial t}\int_0^{k_0}F(k)k^2dk<k_0^2\epsilon, \tag{4}$$

where ϵ equals the total energy supply per unit volume to the fluid and is given as[3]

$$\epsilon=\sqrt{3}k_0v_0^3\kappa/8. \tag{5}$$

On the other hand, the right-hand side of (2) is always greater than

$$2\nu\int_{k_0}^{k_s}F(k)k^4dk\cong\frac{\nu}{5}v_0^2k_0^{3/2}k_s^{10/3}. \tag{6}$$

Hence,

$$\frac{-\frac{\partial}{\partial t}\int_0^\infty F(k)k^2dk}{2\nu\int_0^\infty F(k)k^4dk}\lesssim 56R_0^{-3/2}\kappa^{-3/2}. \tag{7}$$

We can easily see that (7) is in direct contradiction to (2). Since (7) is obtained by using only the behavior of $F(k)$ in the Kolmogoroff region ($F(k)\propto k^{-5/3}$), this contradiction will hold as long as one assumes the existence of a Kolmogoroff region for two-dimensional turbulence.

In the three-dimensional case, the vorticity may be changed by the extension of the vortex filament, and the application of the vorticity equation does not lead to any contradiction.[3]

The author wishes to thank Professor W. Heisenberg for an illuminating communication.

[1] A. N. Kolmogoroff, Compt. rend. acad. sci. U.R.S.S. 30, 301 (1941); 32, 16 (1941); L. Onsager, Nuovo Cimento 6 (9), 279 (1949); C. F. von Weizsäcker, Z. Physik 124, 614 (1948).
[2] W. Heisenberg, Z. Physik 124, 628 (1948).
[3] T. D. Lee, Phys. Rev. 77, 842 (1950).

The Generality of Mixed Gas Flows

MAX M. MUNK
Naval Ordinance Laboratory, Washington, D. C.
(Received January 25, 1951)

W. R. SEARS, in the August, 1950, issue of this Journal, speaks out in favor of neighbors, but still seems to reserve some doubts about the existence of neighbor solutions of mixed flows; that is, of solutions for neighbor contours. The following argument should convince him that neighbors are positively the rule, particularly with plane two-dimensional flows.

These flows are linear in the hodograph and, therefore, admit these superpositions. In the physical plane, the composite flow assigns, to each velocity vector, the center of gravity of the points occupied by the same vector in the constituent flows. Now superpose thus to the mixed flow a weak secondary constituent flow single valued through the pertinent portion of the hodograph. All velocities of the secondary flow taking part in the superposition will be contained within a small area in the physical plane. An immense variety of such secondary flows is available. Each gives rise to a different neighbor solution with the composite contour as close to the initial contour as desired.

The composite flow will be free of limiting lines in general. The discriminant, which becomes zero at the limiting line, is linear in the hodograph. (See reference 8 of Sears.) It is related to the angle between the streamline image and the characteristic line in the hodograph. This discriminant can be supposed to be larger than zero throughout the flow and also through a strip within a certain distance from the contour. The secondary constituent discriminant is finite and small, and by making the secondary constituent smaller and smaller the composite discriminant may be kept from becoming zero.

The composite flow will include an unbroken contour. For not only the legendre reciprocal potential but also the stream function is linear in the hodograph plane. The two impact points, having initially equal velocity and equal stream function, will, therefore, be transformed into two composite impact points having equal stream function.

Decay of Spectral Components in Isotropic Turbulence

HIROSHI SATO
Institute of Industrial Science, University of Tokyo, Chiba, Japan
(Received January 8, 1951)

ALTHOUGH many measurements have been reported for determining the energy spectrum of turbulence,[1,2] there are only a few data about the decay of spectral components. The author observed it by a rather direct method and some interesting results were obtained.

The wind tunnel used is of the single open-jet type, having a rectangular cross section of 60cm × 60cm. A wooden grid is placed at its opening. The dimensions of the grid are as follows: mesh length = 1 in., rod breadth = $\frac{1}{8}$ in., depth = $\frac{1}{2}$ in. The electric circuit for the hot-wire anemometer consists of three elements: the pre-amplifier, the filter, and the power amplifier. The amplifier is an ordinary one. The filter is of the phase-shift band-pass type, Q being about 50. The output current is read by a thermocouple-type milliammeter. A new method of measurement was adopted. We traveled the hot wire longitudinally, fixing the tuning frequency of the filter. The output meter indicated the decay of spectral components. This method has many advantages: the required time is short; if we consider the energy ratio only, we need not know the sensitivity and frequency response of the hot wire and amplifier. Therefore, the accuracy of measurement is high.

Experiments were made for the wind speed $U=5$ m/sec and 10 m/sec. Both results coincide when the relative energy is plotted against the wave number. The decay curve for $U=10$ m/sec is shown in Fig. 1, where x is the distance behind the grid, and E is

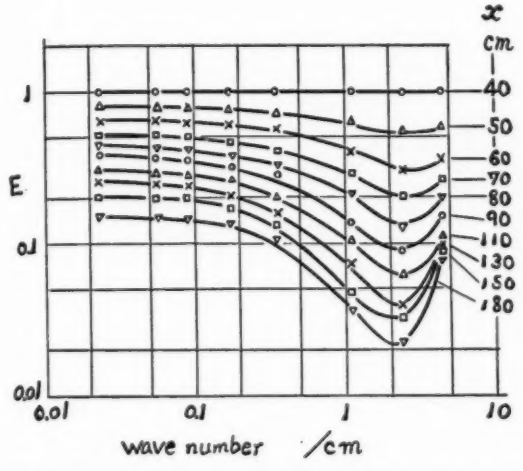


FIG. 1. Decay of spectral components; x, distance behind grid; E, relative energy component.

the relative energy component, taking $x=40$ cm as unity. From this result it may be concluded that (1) the decay law of energy is different for components of various wave numbers; and (2) there exists the component for which the decay is most severe, the corresponding wave number decreasing during the decay.

The result obtained is of the so-called "initial period."[3] The author is now engaged in the more extensive measurements, whose details will be published in the *Journal of the Physical Society of Japan*.

The author would like to acknowledge the valuable advice of Professor I. Tani and Professor H. Kumagai.

[1] H. L. Dryden, Proc. V Intern. Congr. Appl. Mech. 362 (1938).
[2] Liepmann, Laufer, and Liepmann, Final Rep. N.A.C.A. Contract NAW 5442 (1948).
[3] G. K. Batchelor and L. A. Townsend, Proc. Roy. Soc. (London) **A193**, 539 (1948).

Propagation on Dielectric Coated Wires

E. T. KORNHAUSER
Lyman Laboratory, Harvard University, Cambridge, Massachusetts
(Received February 5, 1951)

IN his interesting paper on the theory and application of dielectric coated wires as wave guides, G. Goubau[1] obtains the straightforward solution for the rotationally symmetric E-wave,

$$\frac{J_0(\gamma_i a')N_0(\gamma_i a)-J_0(\gamma_i a)N_0(\gamma_i a')}{J_1(\gamma_i a')N_0(\gamma_i a)-J_0(\gamma_i a)N_1(\gamma_i a')}=\frac{i\gamma'}{\gamma_i}\cdot\frac{\epsilon_i}{\epsilon}\cdot\frac{H_0^{(1)}(i\gamma' a')}{H_1^{(1)}(i\gamma' a')}. \quad (1)$$

In obtaining a useful approximation to this equation, Goubau restricts himself to the case of $\gamma_i a' \ll 1$, so that the left-hand side of (1) becomes $-\gamma_i a' \ln(a'/a)$.

However, it might be pointed out that this logarithmic approximation has a considerably wider applicability than the restriction imposed in obtaining it would imply. For the case $[(a'/a)-1] \ll 1$, Goubau has mentioned that a first-order approximation by Taylor series is possible, yielding $-\gamma_i(a'-a)$ for the left side. It may be noted that this result is in first-order agreement with the previous approximation if, indeed, we have $[(a'/a)-1] \ll 1$.

Moreover, a second-order approximation by Taylor series is possible here, a somewhat similar calculation having been performed by L. Brillouin.[2] In this calculation, terms in $[(a'/a)-1]^3$ are retained, and the second derivatives are eliminated by use of the bessel equation of order zero. The resulting approximation to the left side of (1) is $-\gamma_i \frac{1}{2}(a'+a)[(a'/a)-1]$, which can be seen to agree with the logarithmic approximation through terms in $[(a'/a)-1]^3$. Hence, it appears that the logarithmic approximation is a good one even in the less restrictive case of $[(a'/a)-1]^3 \ll 1$. The accompanying table serves to indicate the usefulness of this approximation in cases in which its original condition is fulfilled, as well as cases in which it is not.

TABLE I. Examples of approximation to

$$F(\gamma a', \gamma a) = -\frac{J_0(\gamma a')N_0(\gamma a)-J_0(\gamma a)N_0(\gamma a')}{J_1(\gamma a')N_0(\gamma a)-J_0(\gamma a)N_1(\gamma a')}.$$

(a) Logarithmic approximation, $\gamma a' \ln(a'/a)$. (b) Second-order Taylor series, $\frac{1}{2}\gamma(a'+a)[(a'/a)-1]$. (c) First-order Taylor series, $\gamma(a'-a)$.

			Approximations		
$\gamma a'$	γa	$F(\gamma a', \gamma a)$	(a)	(b)	(c)
0.02	0.01	0.01386	0.01386	0.01500	0.01
0.1	0.02	0.1617	0.1609	0.2400	0.08
0.1	0.05	0.06939	0.06932	0.07500	0.05
0.1	0.09	0.01053	0.01054	0.01056	0.01
0.5	0.45	0.05273	0.05268	0.05278	0.05
1.0	0.9	0.1057	0.1054	0.1056	0.1
5.1	5.0	0.1013	0.1010	0.1010	0.1

[1] G. Goubau, J. Appl. Phys. **21**, 1119 (1950).
[2] L. Brillouin, *Spiraled Coils as Waveguides* (ONR Technical Report No. 10, Cruft Laboratory, Harvard University, March 10, 1947).

Coefficients of Thermal Expansion of Au-Cd Alloys Containing 47.5 Atomic Percent Cd

LO-CHING CHANG
*Columbia University, New York, New York**
(Received January 31, 1951)

THE coefficients of thermal expansion of single crystals of Au–Cd β_1 alloy (CsCl structure, two atoms per unit cell) and β' alloy (orthorhombic structure, four atoms per unit cell), containing 47.5 atomic percent Cd, were determined by means of x-ray single-crystal oscillation method. The β_1 alloy undergoes a diffusionless transformation to β' on cooling to about 60°C, and the reverse diffusionless transformation from β' to β_1 takes place on heating the β' alloy to about 70–80°C.[1]

The experimental procedures used are: First, rotate the desired

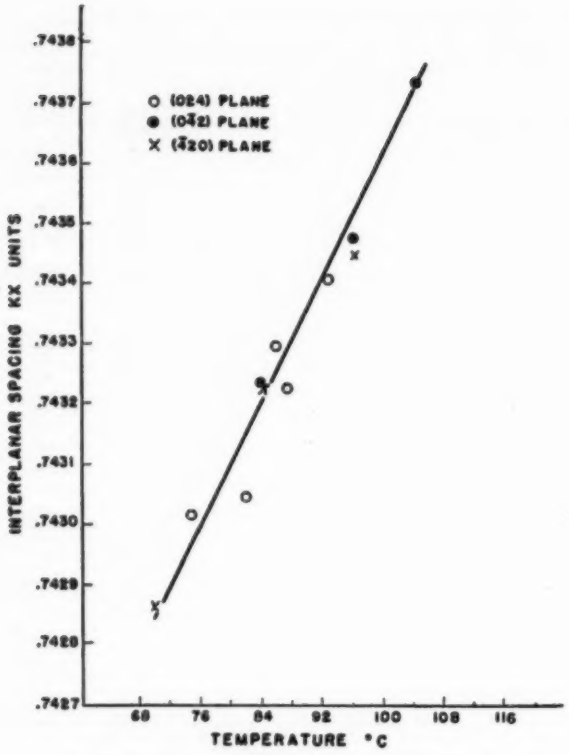


FIG. 1. Spacings *versus* temperature of (420) type plane of β_1 alloy (crystal No. 12).

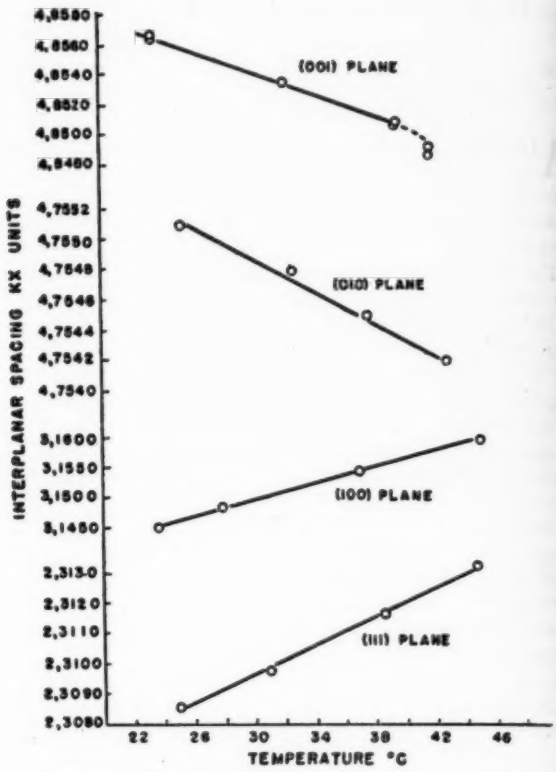


FIG. 2. Spacings *versus* temperature of various planes of β' alloy (crystal No. 13).

plane of the single-crystal specimen (mounted on a goniometer) into position for high angle reflection ($\theta=75°$ or higher) by x-ray characteristic radiation of accurately known wavelength. Second, control the specimen temperature to within $\pm\frac{1}{2}°$ of the desired value, oscillate the specimen about $2\frac{1}{2}°$ each way, from the reflecting position, and obtain an exposure. Third, rotate the camera 180° about the axis of the x-ray beam and obtain a second exposure. Thus, the film will have two diametrically opposite spots, the distance between which, $2l$, is related to the film specimen distance, L; the interplanar spacing, d; the Bragg angle, θ; and the wavelength of x-ray, λ; by the following equations:

$$l/L=\tan(180°-2\theta), \quad (1)$$

$$\lambda=2d\sin\theta. \quad (2)$$

Knowing l, L, and λ, the interplanar spacing d at any given temperature can be calculated from Eqs. (1) and (2). The interplanar spacings are thus determined at about 5°C intervals. This method, although rather laborious, has the advantage of measuring directly the coefficient of thermal expansion of a single plane of a single crystal, over the conventional powder method which yields the average value of the coefficients of expansion of many crystals. The accuracy of measurements depends very much on the Bragg angle. The Bragg angles used in the present study varied from 72° to 88°. The over-all accuracy of the determination of coefficients of thermal expansion is about ±3 percent.

The results, plotted in Figs. 1 and 2, yield the following coefficients of thermal expansion of the β_1 and β' alloy: β_1 alloy (cubic) 33×10^{-6} per °C (70–105°C). β' alloy (orthorhombic), directions perpendicular to:

(100) plane	216×10^{-6} per °C	(20–45°C).
(010) plane	-12×10^{-6} per °C	
(001) plane	-74×10^{-6} per °C	
(111) plane	99×10^{-6} per °C	

* This work was assisted by the AEC.

[1] L. C. Chang and T. A. Read, J. Metals (Japan) **191**, 47 (1951).